This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

USA

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, 1/70, C12P 19/34, C07H 21/02, 21/04, G01N 33/53, C07K 15/28, 5/00

A1

(11) International Publication Number: **WO 97/17470**

(43) International Publication Date: 15 May 1997 (15.05.97)

(21) International Application Number: PCT/US96/17877

(22) International Filing Date: 8 November 1996 (08.11.96)

(30) Priority Data:
08/555,394 9 November 1995 (09.11.95) US

(71)(72) Applicant and Inventor: HOLLAND, James, F. [US/US]; 31 Mamaroneck Road, Scarsdale, NY 10583 (US).

(72) Inventor: POGO, Beatriz, G., T.; 237 Nyac Avenue, Pelham, NY 10803 (US).

(74) Agents: CLARK, Richard, S. et al.; Brumbaugh, Graves, Donohue & Raymond, 30 Rockefeller Plaza, New York, NY 10112 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(57) Abstract

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

BNSDOCID: <WO 9717470A1>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

BNSDOCID: <WO 9717470A1>

Description

Detection Of Mammary Tumor Virus-Like Sequences In Human Breast Cancer

Cross-Reference to Related Application

This application is a continuation-in-part application of U.S. Serial No. 08/555,394, filed November 9, 1995.

Statement Regarding Federally Sponsored Research

This invention was made with funds from the U.S. government, which has certain rights in the invention.

Introduction

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

Background of the Invention

A large body of information has accumulated about the molecular biology of MMTV (reviewed in Slagle, B.L. et al., 1987, in "Cellular and Molecular Biology of Mammary Cancer", Kidwell et al., eds., Plenum Press, NY. pp 275-306). Mouse mammary tumor virus (MMTV) is associated with a high incidence of breast cancer in certain strains of mice (over 90% among females), and has been regarded as a potential model for human disease.

The MMTV virus does not carry a transforming oncogene, but rather acts as an insertional mutagen with several proviral insertion loci designated int-1

or wnt-1 (Nusse R. et al., 1982, Cell 31:99-109) int-2 (Peters, G. et al., 1983, Cell 33:369-377) int-3 (Gallahan, D. et al., 1987, J. Virol. 61:218-220) int-4 (Roelink, H. et al., 1990, Proc. Natl. acad. Sci. USA 87:4519-4523) and int-5 (Morris, V.L., et al. 1991, Oncogene Research 6:53-63), which encode for growth factors or other related proteins. These genes are not expressed in normal mammary tissue but become activated after integration of MMTV provirus into the adjacent chromosomal DNA.

The human homolog of the int-2 locus has been located on chromosome 11 (Casey, G. et al., 1986, Mol. Cell Biol. 6:502-510) and has been found amplified (in 15% of the breast cancers) and also expressed (Lidereau, R. et al., 1988, Oncogene Res 2:285-291; Zhou, D.J. et al., 1988, Oncogene 2:279-282; Liscia, D.S. et al., 1989, Oncogene 4:1219-1224; Meyers, S.L. et al., 1990, Cancer Res 50:5911-5918). It may be significant that in tumors from Parsi women, who have a high incidence of breast tumors, the int-2 locus is amplified in 50% of the cases (Barnabas-Sohi, N. et al., 1993, Breast Dis. 6:13-26). The amplification of int-2 and other genes in 11q13 is indicative of poor prognosis (Schuwring, E. et al., 1992, Cancer Research 52:5229-5234; Champeme, M-H, et al., 1995, Genes, Chromosomes and Cancer 12:128-133). Both mouse and human int-2 have been sequenced (Moore, R. et al., 1986, EMBO J 5:919-924). The gene encodes a protein of about 27 kilodaltons (KD) which shows homology to both basic and acidic fibroblast growth factors (Dickson, C. et al. 1987, Nature (London) 326:833).

However, efforts to demonstrate the presence of viruses in human breast cancer through search for viral particles, immunological cross-reactivity, or sequence homology have yielded contradictory results. Detectable MMTV env gene-related antigenic reactivity has been found in tissue sections of breast cancer

BNSDOCID: <WO 9717470A1>

(Mesa-Tejada et al., 1978, Proc. Natl. Acad. Sci. USA 75:1529-1533; Levine, P. et al., 1980, Proc. Am. Assoc. Cancer Res. 21:170; Lloyd, R. et al., 1983, Cancer 51:654-661), breast cancer cells in culture (Litvinov, S.V. and Golovkina, T.V., 1989, Acta Virologica 33:137-142), human milk (Zotter S. et al., 1980, Eur. J. Cancer 16:455-467) in sera of patients (Day, N.K. et al., 1981, Proc. Natl. Acad. Sci. USA 78:2483-2487), in cyst fluid (Witkin, S.S. et al., 1981, J. Clin. Invest. 67:216-222) and in particles produced by a human breast carcinoma cell line (Keydar, I. et al., 1984, Proc. Natl. Acad. Sci. USA 81:4188-4192). Sequence homology to MMTV has been found in human DNA under low stringency conditions of hybridization (Callahan, R. et al., 1982, Proc. Natl. Acad. Sci. USA 79:5503-5507) and RNA related to MMTV has been detected in human breast cancer cells (Axel, R. et al., 1972, Nature 235:32-36). The presence of MMTV related sequences in lymphocytes from patients with breast cancer has been reported (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331), as well as detection of reverse transcriptase (RT) activity in their monocytes (Al-Sumidaie, A.M. et al., 1988, Lancet 1:5-8). May and Westley (May and Westley, 1989, Cancer Research 49:3879-3883) have reported the presence of MMTV-like sequences arranged as tandem repeats only in DNA from breast cancer cells.

These results have been difficult to interpret, and theories linking MMTV or a related virus with human breast cancer have fallen out of favor, in view of the relatively recent discovery of human endogenous retroviral sequences ("HERs"; Westley, B. et al., 1986, J. Virol. 60:743-749; Ono, M. et al., 1986, J. Virol. 60:589-598; Faff, O. et al., 1992, J. Gen. Virology 73:1087-1097). Data which could be interpreted to demonstrate the presence of MMTV-related sequences could be more readily explained by endogenous human

BNSDOCID: <WO 9717470A1>

retroviral sequences. Adding further confusion to the picture, env-gene related antigenicity has been detected in epitopes of human proteins (Hareuveni, M. et al., 1990, Int. J. Cancer 46:1134-1135).

Brief Summary of the Invention

The present invention relates to methods for diagnosing breast cancer in humans in which the presence of mouse mammary tumor virus env gene-like sequences bears a positive correlation to the existence of malignant breast disease. It is based, at least in part, on the discovery that 38 to 40 percent of human breast cancer tissue samples tested contained gene sequences homologous to the mouse mammary tumor virus env gene that are substantially absent from other human tumors and tissues. The invention also relates to methods for diagnosing breast caner in humans in which the presence of retrovirus proviral fragments substantially homologous to the env gene and/or 3' LTR sequence of MMTV are detected. The molecular probes used in these experiments were designed to avoid cross-hybridization with endogenous human retroviral sequences. The present invention further provides for compositions of molecular probes which may be utilized in such diagnostic methods.

Brief Description of the Figures

FIGURE 1: Amplification of 660 bp of MMTV-like env gene. DNA was extracted from frozen tissues. PCR was performed using primers 1 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'^{32}P-end-labeled probe 2. Lanes 1 and 3: breast cancer; lanes 2 and 4: normal breast; lane 5: control reaction (no DNA); lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 510 bp band.

FIGURE 2: Nested PCR. A: 2% agarose gel electrophoresis. 1: Amplification of 686 bp of MMTV-like env

BNSDOCID: <WO 9717470A1>

gene sequences using primers 1 and 4 and the product of reaction A 1 as template. 2: Amplification of 250 bp of MMTV-like env gene sequences using primers 2 and 3. B, 1 and 2: Southern blot hybridization of the amplified products using probe 5'-^{32}P end-labeled probe 2a.

FIGURE 3: Amplification of 250 bp of MMTV-like env gene. DNA was extracted from paraffin-embedded tissue sections. PCR was performed using primers 2 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'-^{32}P-labeled probe 2a. Lane 1: normal breast; lanes 2 to 5: breast cancer; lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 298 bp band.

FIGURE 4: Nucleotide sequence of the cloned MMTV env gene-like sequences as compared to the env sequences of the GR and BR6 strains of MMTV using the GCG program. *:potential glycosylation site, |:mismatch to MMTV.

FIGURE 5: Southern blot hybridization of genomic DNA. DNA was extracted from frozen tissues or cell lines, digested with EcoR1 and transferred to nitrocellulose paper. Hybridization with ^{32}P-labeled clone 166. DNA from A, B, and G: env gene positive breast cancer; C and D: env negative breast cancer; E and F: normal breast; H:MCF-7 cells. M: molecular weight marker, Arrow indicates 9kb band.

FIGURE 6: Southern blot hybridization of genomic DNA. Experimental conditions as in Fig. 5. DNA from A and B: env negative breast cancer; C and D: env positive breast cancer; E: molecular weight marker (non-labelled); F. to H: normal breast. Arrow indicates position of 9 kb marker.

FIGURE 7: Map of MMTV.

FIGURE 8: Comparison of the nucleic acid sequence of mouse mammary tumor env gene ("MMTENV"), showing residues 976-1640, with the nucleic acid sequence of a

BNSDOCID: <WO 9717470A1>

representative 660 bp sequence obtained by PCR reaction of DNA from human breast cancer tissue ("MS1627").

FIGURE 9: Sequence of an about 2.6 kb MMTV-like fragment detected in a human breast carcinoma.

Detailed Description of the Invention

The present invention relates to methods and compositions for diagnosing breast cancer in humans.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which (i) hybridizes to a gene of mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. A "gene of mouse mammary tumor virus" includes, but is not limited to, the gag, pol, and env genes and the 5′ LTR and 3′ LTR sequences of MMTV. In preferred embodiments of the invention, the mouse mammary tumor virus (hereafter "MMTV") gene is the env gene and/or the 3′ LTR sequence. The term "hybridize" is used to refer to routine DNA-DNA or DNA-RNA hybridization techniques under what would be regarded, by the skilled artisan, as stringent hybridization conditions. The phrase "is present" indicates that a native form of the molecule, in an unpurified state (for example, as part of chromosomal DNA), may be detected by a standard laboratory technique, such as Southern blot or polymerase chain reaction (PCR). To be "present", the molecule may be detectable by one technique but not others. To be present in "less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects", all non-breast cancer tissue samples are considered together, but the total number of samples must be large enough to give the 5 percent

BNSDOCID: <WO 9717470A1>

value statistical significance that would be reasonable to the skilled artisan.

In order to identify such a nucleic acid molecule, the sequence of MMTV may be compared, using a computer database, to known human DNA sequences, and portions of MMTV which are less than or equal to 25 percent homologous to a human sequence may be selected for further study. The term "homologous", as used herein, refers to the presence of identical residues; for example, a first sequence is considered 25 percent homologous to a second sequence if it shares 25 percent of the residues of the first sequence. Since there is relatively greater likelihood that MMTV may bear similarity to human retroviral-like sequences, it may be preferable to evaluate whether a particular MMTV nucleic acid sequence is homologous to such sequences, for example, as endogenous human retrovirus sequences. A prototype of such viruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598).

Once an MMTV gene sequence which is less than or equal to 25 percent homologous to a human DNA sequence, such as a human endogenous retroviral sequence, is identified, the presence of nucleic acid molecules having the MMTV gene sequence in human breast cancer tissues and other tissues may be evaluated. Such evaluations may be performed either by Southern blot techniques, or, preferably, by polymerase chain reaction (PCR) techniques, which are more sensitive. In such a way, MMTV gene sequences which (i) hybridize to at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects and (ii) hybridize to less than 5 percent of DNA samples prepared from human tissues other than breast cancer tissues may be identified. A nucleic acid molecule having a MMTV gene sequence which satisfies these requirements may then be used in diagnostic methods which detect the presence of such sequence in human

BNSDOCID: <WO 9717470A1>

breast tissue by standard techniques, including PCR techniques which assay for the presence of the molecule, but also, where appropriate, Southern blot, Northern blot, or Western blot techniques, to name but a few.

In preferred embodiments, the present invention relates to a portion of MMTV localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; see Fig. 7). This about 660 bp sequence (hereafter, "the 660 bp sequence") has been found to exhibit low (16 percent) homology to the prototype human endogenous retrovirus HERV-K10, using the IBI/Pustell Sequence Analysis Program, and has also been shown to be present in 121 (38.5%) of 314 unselected breast cancer tissue samples, in cultured breast cancer cells, in 2 of 29 breast fibroadenomas (6.9%) and in 2 of 107 breast specimens from reduction mammoplasties (1.8%). The sequence was not found in normal tissues including breast, lymphocytes from breast cancer patients nor in other human cancers or cell lines (see example section, infra). Similarly, an about 250 bp sequence (hereafter "the 250 bp sequence"), between positions 1388 and 1640 in the env gene, and therefore falling within the 660 bp sequence, was detected in 60 (39.7%) of 151 breast cancer, and in one of 27 normal breast samples assayed from paraffin-embedded sections. Cloning and sequencing of the 660 bp and 250 bp sequences demonstrated that they are 95-99% homologous to MMTV env gene, but not to the known human endogenous retroviruses ("HERs") nor to other viral or human genes (<18%).

In another preferred embodiment, the present invention relates to a a nucleic acid molecule which corresponds to a retroviral genomic fragment which has substantial homology to 3' LTR and/or env gene of the MMTV genome, and is found in a substantial percentage of breast cancer samples. By substantial percentage is

BNSDOCID: <WO 9717470A1>

meant at least 20% of tested breast cancer samples. Such a sequence is preferably comprised of the 3' LTR region and all or part of the env gene, although it may include more sequences of a retroviral genome. Most preferably, the sequence is at least comprised of an about 2.6 kb fragment which comprises the 1,228 base pair (bp) sequence of the 3' LTR sequence and 1,336 bp of the env gene sequence of MMTV (Fig. 9) (SEQ ID NO:20). When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

Retrovirus proviral sequences can be detected by PCR technology using primers derived from the MMTV genome. Such primers include primer 5L, containing the nucleotides 7376-7395 of the MMTV BR6 genome (5'-3': CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3, containing nucleotides 9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Other primers which correspond to or are homologous to MMTV sequences can be used as primers. Nucleotide fragments which correspond to or are homologous to the retroviral sequences isolated from the breast cancer samples can also be used to amplify additional retroviral fragments from the samples. Long PCR techniques can be used to amplify longer stretches of a proviral sequence.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which hybridizes to the about 2.6 kb retroviral fragment shown in Fig. 9 under stringent conditions or is at least 90 percent homologous to said fragment using the MacVector homology determining program which may be used to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods.

BNSDOCID: <WO 9717470A1>

Nucleic acids having the 660 bp sequence, the 250 bp sequence, or all or part of the about 2.6 kb sequence, may therefore be used, according to the invention, to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods. Where PCR methods are used, primers such as those listed in Table 1, below, may be utilized.

The present invention provides for compositions comprising essentially purified and isolated nucleic acid having the 660 bp sequence or the 250 bp sequence or an at least five bp, and preferably greater than or equal to ten bp, subsequence thereof. In order to maintain the desired specificity, such nucleic acid molecules may preferably contain sequence falling within the 660 bp sequence, but preferably do not contain sequences from other portions of the MMTV genome, which may, undesirably, hybridize to human sequences which are not breast cancer specific, such as HERs. Accordingly, the present invention provides for compositions wherein the isolated and purified nucleic acid molecule comprises at least a portion having a nucleic acid sequence which hybridizes to a region of the mouse mammary tumor virus <u>env</u> gene between residues 976 and 1640, or between residues 1388 and 1640, and wherein the isolated and purified nucleic acid molecule does not hybridize to any other region of the MMTV genome.

The 660 bp sequence, in various embodiments, may have a number of nucleotide sequences. For example, in one embodiment, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MMTENV-like sequence" (SEQ ID NO:17), which depicts the MMTV <u>env</u> sequence between residues 976 and 1640. In a second series of embodiments, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MS1627" (SEQ ID NO:18), which depicts a predominant sequence for the 660 bp sequence as it has been defined by

BNSDOCID: <WO 9717470A1>

sequencing analysis of the products of PCR reactions using DNA from human breast cancer tissues. In still further embodiments, the 660 bp sequence may have various other nucleotide sequences obtained by sequencing the results of PCR reactions to detect the presence of 660 bp sequence in human breast cancer tissues.

In related embodiments, the present invention provides for compositions comprising PCR primers that may be used to detect the presence of the forementioned molecules or other MMTV-like sequences. For example, the compositions may comprise one or more of the following primer molecules (5′ - 3′): CCTCACTGCCAGATC (SEQ ID NO:1); GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2); CCTCACTGCCAGATCGCCT (SEQ ID NO:3); TACATCTGCCTGTGTTAC (SEQ ID NO:4); CCTACATCTGCCTGTGTTAC (SEQ ID NO:5); CCGCCATACGTGCTG (SEQ ID NO:6); ATCTGTGGCATACCT (SEQ ID NO:7); GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8); ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); TACAGGTAGCAGCACGTATG (SEQ ID NO:12); CGAACAGACACAAACACACG (SEQ ID NO:19).

The use of such compositions and molecules in PCR and Southern blot techniques is illustrated in the non-limiting examples set forth below. The correlation between the presence of the MMTV-related nucleic acid molecules described above and breast cancer allows such molecules and compositions to be utilized in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer, wherein the detection of such nucleic acid molecules bears a positive correlation to the existence of breast cancer in a human. The results of such evaluation, together with additional clinical symptoms, signs, and laboratory test values, may be used to formulate the complete diagnosis of the patient.

BNSDOCID: <WO 9717470A1>

In further related embodiments, the present invention provides for an essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. In preferred embodiments, the MMTV gene is the env gene.

Such peptides may be used in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

The present invention also provides for antibodies (including monoclonal and polyclonal) antibodies which specifically bind to such peptides. Such antibodies may be used in methods of diagnosing breast cancer, for example, but not by way of limitation, by Western blot, immunofluorescent techniques, and so forth.

In nonlimiting embodiments of the invention, the skilled artisan may evaluate MMTV-like nucleic acid molecules for regions which would be considered likely to encode immunogenic peptides (using, for example, hydropathy plots). Such peptides may then be sequenced and used to produce antibodies that may be employed in diagnostic methods as set forth above.

For example, certain peptides encoded by portions of the 660 bp sequence have been synthesized. These

BNSDOCID: <WO 9717470A1>

peptides, which have the sequences LKRPGFQEHEMI (SEQ ID NO:13) and GLPHLIDIEKRG (SEQ ID NO:14), have been used to produce antibodies in rabbits, and the resulting antisera have successfully identified breast cancer cells positive for MMTV env-like sequences by PCR assay. Other peptides encoded by 660 bp sequence which may be useful according to the invention include TNCLDSSAYDTA (SEQ ID NO:15) and DIGDEPWFDD (SEQ ID NO:16).

6. Example: The Detection of Mouse Mammary Tumor Virus Env Gene-Like Sequences in Human Breast Cancer Cells and Tissues

6.1. Materials and Methods

DNA from breast cancer tissue and other human cancer tissues, human placentas, normal human tissues including breast, and from several human cell lines (including eight breast cancer cell lines), and two normal breast cell lines was extracted following the procedure of Delli Bovi et al. (1986, Cancer Res. 46:6333-6338). The DNA was resuspended in a solution containing 0.05 M Tris HCl buffer, pH 7.8, and 0.1 mM EDTA, and the amount of DNA recovered was determined by microfluorometry using Hoechst 33258 dye (Cesarone, C. et al., 1979, Anal Biochem 100:188-197). Plasmids containing the cloned genes of MMTV were obtained from the ATCC, propagated in Escherichia coli cultures and purified using anion-exchange minicolumns (Qiagen) or by precipitation with polyethylene glycol (Sambrook J., et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). Oligonucleotide primers were synthesized at the core facilities of the Brookdale Molecular Biology Center at Mount Sinai School of Medicine.

Polymerase chain reaction (PCR) was performed using Taq polymerase following the conditions recommended by the manufacturer (Perkin Elmer Cetus)

BNSDOCID: <WO 9717470A1>

with regard to buffer, Mg^{2+} and nucleotide concentrations. Thermocycling was performed in a DNA cycler by denaturation at 94° C for 3 min. followed by either 35 or 50 cycles of 94°C for 1.5 min., 50° C for 2 min. and 72°C for 3 min. The ability of the PCR to amplify the selected regions of the MMTV env gene was tested by using as positive templates the cloned MMTV env gene and the genomic DNA of the MCF-7 cell line, since it was shown to express gp52 immunological determinants (Yang, N.S., et al., 1975, J. Natl. Cancer Inst. 61:1205-1208). Optimal Mg^{2+}, primer concentrations and requirements for the different cycling temperatures were determined with these templates. The master mix as recommended by the manufacturer was used. To detect possible contamination of the master mix components, a reaction without template was routinely tested. γ DNA and control primers provided by the manufacturer were used as control for polymerase activity. As an internal control, amplification of a 120 bp sequence estrogen receptor gene was assayed using primers designed and generously provided by Dr. Beth Schachter, (Mount Sinai School of Medicine, N.Y.). In addition, primers for actin 5 gene amplification were also used.

The product of the PCR was analyzed by electrophoresis in a 2% agarose gel. A 1 kb DNA ladder (Gibco BRL) was used to identify the size of the PCR product. To determine if the amplified sequences of the middle region of the 660 bp faithfully reproduced the sequences of the env gene of MMTV, an 18-mer sequence within the env gene was used as a probe for the 660 bp amplified sequence. The 18-mer probe was 5′ end-labeled with ^{32}P-ATP using T4 polynucleotide kinase and purified by the NENSORB nucleic acid purification cartridge (NEN). Southern blot hybridization was performed using the conditions described by (Saiki et al.,1985, Science 230:1350-1354).

BNSDOCID: <WO 9717470A1>

The product of the PCR (660 bp or 250 bp) was cloned directly from the reaction mixture into the TA cloning vector (Invitrogen) using the TA cloning kit and following the conditions recommended by the supplier. Direct cloning of the fragment isolated from the gel, was also performed. Plasmid DNA was purified by CsCl density gradient centrifugation or by precipitation with polyethylene glycol (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor), restricted with HindIII and EcoRl, electrophoresed in 2% agarose gels and transferred to nitrocellulose filters. Southern blot hybridization was carried out using a 5'-terminal labeled internal probe as described above. Cloning procedures were performed in laboratories totally separate from those where PCR was carried out. Automated DNA sequencing (using Applied Technology Sequencer Model 373A) was performed in the Brookdale Molecular Biology Center. Sequence homology was determined using the IBI MacVector GenBank and GCG Programs.

To prevent contamination of the samples, processing of human tissues was performed in a laminar flow hood. DNA extractions were done in a chemical hood located in a different room from that where PCR was performed. PCR assays were assembled in a biological hood provided with ultraviolet light. Aerosol resistant tips and dedicated positive-displacement pipettes were used throughout. All equipment used for PCR (microcentrifuge, electrophoresis apparatus, pipettors) was cleaned each time with 10% sodium hypochlorite to assure DNA decontamination (Prince and Andrus, 1992, Biotechniques 12:358-36). After the initial experiments were performed, the plasmid containing the MMTV <u>env</u> gene was frozen and never used again, to avoid contamination. However, to detect plasmid contamination from our own <u>env</u> gene clones,

BNSDOCID: <WO 9717470A1>

primers were designed to amplify plasmid sequences. All the authentic MMTV env positive samples were then tested and found negative for plasmid contamination.

Southern blotting and hybridization were performed as described (Southern, E.M., 1975, J. Mol. Biol. 98:503-517), using the 660 bp cloned sequences labeled by the random primer procedure (Feinberg, A.P., et al., 1983, Anal. Biochem. 132:6-13). Prehybridization and hybridization were performed in a solution containing 6 x SSPE, 5% Denhardt's, 0.5% SDS, 50% formamide, 100 μg/ml denaturated salmon testis DNA, incubated for 18 hrs at 42°C, followed by washings with 2 x SSC and 0.5% SDS at room temperature and at 37°C and finally in 0.1 x SSC with 0.5% SDS at 68°C for 30 min (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). For paraffin-embedded tissue sections the conditions described by Wright and Manos (1990, in "PCR Protocols", Innis et al., eds., Academic Press, pp. 153-158) were followed using primers designed to detect a 250 bp sequence.

6.2. Results

6.2.1. Selection of Specific MMTV Env Gene Sequences

A computer search for MMTV env gene homologous sequences was first performed, since sequence homology between the human endogenous retroviral sequences and MMTV had been described. The prototype of this group of human endogenous retroviruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598). The sequences of the env gene of MMTV (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504) were aligned with sequences of the env gene of the human endogenous retrovirus HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598), using the IBI/Pustell Sequence Analysis Program. A region of 660 bp of low homology (16%) was localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504). This

BNSDOCID: <WO 9717470A1>

internal domain of the outer membrane of the env gene has only one glycosylation site and is highly conserved between strains. Two primers comprising 15 bp sequences at positions 976-990 (primer 1) and 1626-1640 (primer 3) were first synthesized. Later longer primers were synthesized (1N and 3N). An 18-mer sequence in the middle of the 660 bp MMTV env region (1388-1405) (primer 2) was used as a probe to identify the 660 bp sequence. A second oligomer probe was synthesized comprising the sequence 1554 to 1568 (primer 2a) to be used for hybridization when a sequence of around 250 bp (between positions 1388 and 1640) was amplified. For nested PCR reactions (Mullis, K.B. and Faloona, F.A., 1987, Meth Enzymol 155:335-350), another primer comprising sequences 1647 to 1661 (primer 4) was synthesized to be used with primer 1 in the first reaction and primers 2 and 3 in the second. Modified primers with GC clamps and extra sequences were also synthesized and used in the PCR (primers 1a and 3a). Another set of primers comprising sequences 974 to 1003 (5L) and 1558 to 1577 (3L) were subsequently developed because their Tm's matched and provided better amplification than the original primers. The sequences are represented in Table 1. All of them were productive in amplification reactions.

BNSDOCID: <WO 9717470A1>

Table 1. Primer and probe sequences and location in murine mammary tumor virus env gene

Designation	Sequence (5'-3')	Location
1	CCTCACTGCCAGATC	976-990
1a	GGGAATTCCTCACTGCCAGATC	976-990
1N	CCTCACTGCCAGATCGCCT	976-993
2	TACATCTGCCTGTGTTAC	1388-1405
2N	CCTACATCTGCCTGTGTTAC	1386-1405
2a	CCGCCATACGTGCTG	1554-1568
3	ATCTGTGGCATACCT	1640-1626
3a	GGGAATTCATCTGTGGCATACCT	1640-1626
3N	ATCTGTGGCATACCTAAAGG	1640-1621
4	GAATCGCTTGGCTCG	1661-1647
5L	CCAGATCGCCTTTAAGAAGG	984-1003
3L	TACAGGTAGCAGCACGTATG	1558-1577

6.2.2. Detection of MMTV-Like Env Gene Sequences in Human Breast Tumor DNA

PCR was performed on DNA extracted from breast cancer tissues, normal breast tissues and from the plasmid containing the env gene of MMTV, using primers 1 and 3. Photographs of the ethidium bromide stained gels of the PCR product reveal the presence of an approximately 660 bp sequence in some of the tumors, (Fig. 1A, lanes 1 and 3) but not in the normal tissue samples (Fig. 1A, lanes 2 and 4). As a positive control the MMTV env gene was also amplified (Fig. 1A, lane E). Similar results were obtained with modified primers 1a, 3a, 3L and 5L. Southern blot hybridization of the gel with ^{32}P-labeled 18-mer oligonucleotide (primer 2) indicated that this internal sequence was present in the amplified material (Fig. 1B) and that the bands in the gel were not artifactual.

Our initial effort was to analyze a representative sample of breast cancer specimens as well as normal

BNSDOCID: <WO 9717470A1>

tissues and other tumors. To date 343 breast tumors have been processed, DNA extracted and PCR preformed. Of these 343 tumors, 314 were carcinomas and 29 were fibroadenomas. Amplification of sequences of 660 bp was observed in 121 of the carcinomas (38.5%) and in 2 of the 29 fibroadenomas (6.9%). These sequences were confirmed to be MMTV env gene-like sequences by hybridization with the labeled specific probe containing the internal sequences. These sequences were not detected in the DNAs extracted from 20 normal organs, 23 cancers from other organs and 26 samples of blood lymphocytes including 7 from breast cancer patients whose breast specimens were positive. From 107 samples of normal breast obtained from reduction mammoplasties, 2 were positive (1.8%). In addition to DNA from lymphocytes from seven positive patients, DNA from their normal breast tissue of the operated breast was tested in 4 cases. All were negative (Table 2). Finally, DNA of the MCF-7, and ED (a cell line developed in our laboratory from the pleural effusion of a patient with an env -positive breast tumor) breast cancer cell lines were shown to contain the 660 bp MMTV env gene-like sequences (Table 3), while four other breast cancer cell lines were positive only for the 250 bp sequence (T47-D, BT-474, BT-20 and MDA-MB-231).

BNSDOCID: <WO 9717470A1>

Table 2. Det ction of MMTV env gene-lik s quences in human DNA xtracted from fresh or frozen tissues

Sample	Number	MMTV env gene sequences	% Positive
Breast Carcinomas	314	121	38.5%
Breast Fibroadenomas	29	2	6.9%
Normal Breasts	107	2	1.8%
*Normal Breasts	4	negative	
Tumors other than breast	23	negative	
Normal tissues	20	negative	
Lymphocytes	26	negative	
**Lymphocytes	7	negative	

* Histologically normal tissue from same breast as positive cancer.

** Lymphocytes from breast cancer patients who were positive for MMTV env gene sequences in the tumor.

BNSDOCID: <WO 9717470A1>

Table 3. Detection of MMTV env gene-like sequences in DNA from human cell lines in culture

Human Cell Lines		MMTV env gene sequence
MC-7	(breast carcinoma)	positive
T47-D	" "	negative
BT-20	" "	negative
MDA-MB-231	" "	negative
ZR-75-1	" "	negative
SK-BR 3	" "	negative
BT474	" "	negative
ED	" "	positive
MCF-10	(normal breast)	negative
HB-447	" "	negative
HL-60	(promyelocytic leukemia)	negative
K562	(erythroleukemia)	negative
Jurkat	(T cell leukemia)	negative
Hep 6-2	(hepatoma)	negative

The nested polymerase reaction was used in several instances to increase sensitivity and specificity, thus reducing the probability of false positives. In Fig. 2, results of a representative nested reaction are shown using primers 1 and 4 in the first reaction (Fig. 2A) and 2 and 3 for the 2nd reaction. The specificity of the reaction can be seen in the 2nd amplification (Fig. 2B).

To study a large number of samples and to be able to perform archival studies, PCR of paraffin-embedded tissue sections was also carried out. Primers 2 and 3 were used to amplify a 250 bp sequence within the 660 bp stretch when DNA was extracted from paraffin-embedded tissue sections since larger size sequences are difficult to amplify after fixation. Tumor DNA was amplified (Fig. 3A, lanes 2-5) whereas normal breast DNA was not (Fig. 3A, lane 1). The identification of

BNSDOCID: <WO 9717470A1>

this 250 bp sequence with the MMTV-like env gene was confirmed by hybridization with an internal probe (primer 2a) as shown in Fig. 3B. Using this procedure we have analyzed 151 breast cancer samples and found that 60 (39.7%) possess the 250 bp sequence. Of the 27 normal breast samples obtained from reduction mammoplasties assayed by this procedure, one was positive (3.7%). These results, in conjunction with those obtained from lymphocytes and from normal breast tissue of patients whose breast cancer was PCR positive, indicate that MMTV-like sequences are present in a significant number of human breast cancer DNA which cannot be explained by DNA polymorphism.

6.2.3. Cloning and Sequencing of the MMTV-Like Env Gene Sequences

To find out whether there was homology to MMTV env gene throughout the whole 660 bp stretch, the product of the PCR from 8 different tumors was cloned and sequenced. In Fig. 4 the sequence of different clones comprising around 600 bp are represented, as aligned to the MMTV env gene sequence of the GR and BR6 strains (Redmon, S. and Dickson, C., 1983, EMBO J. 2:125-131). This domain of the env gene in the GR strain is 100% homologous to the C_3H strain and 98% to the BR6 strain (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; Moore, R. et al., 1987, J. Virol. 61:480-490). Evaluation of the clones indicated that homology to MMTV env gene varied from 95% to 99%. Another seven clones comprising only 250 bp were also sequenced. Homology to MMTV env gene varied from 95% to 99% (data not shown). When compared to the human endogenous provirus HERV-K10, the homology of all the clones was less than 15%. When compared against all known viral and human genes (more than 130,000 entries) using the 1B1 MacVector GenBank and GCG programs, the highest homology recorded was 18%.

BNSDOCID: <WO 9717470A1>

6.2.4. Southern Blot Analysis Using Cloned Sequences

To investigate whether the env gene-like sequences were present in human DNA, Southern blot hybridization was performed using the cloned sequence as probe. DNAs from normal breast tissues, env positive or negative breast tumors, tumors other than breast and breast cancer cell lines were restricted with EcoRI and in some instances with PstI, BglII or KpnI. EcoRI is a frequent cutter restriction enzyme that digests MMTV proviral DNA between env and pol genes. Four different cloned 660 bp sequences were used as probes after labeling with ^{32}P by random prime-labeling. Results of some of the Southern blot hybridization experiments are shown in Fig. 5. They reveal the presence of a labeled restriction fragment migrating at approximately 7-8 kb in breast cancer DNA, in ED and two fragments in MCF-7 cells. Different restriction patterns were observed with the other three enzymes. The 660 bp sequence was absent in 10 normal tissues, 10 fibroadenomas and 10 tumors from other tissues. It is important to emphasize that hybridization conditions for these experiments were stringent (as described in Section 6.1) to avoid interference with endogenous sequences that might interact with the probes.

7. Example: Detection of a Retrovirus Proviral Fragment in Human Breast Cancer Cells and Tissues

7.1. Materials and Methods

To detect longer retrovirus proviral fragments in breast cancer samples, DNA was extracted from breast cancer carcinoma tissue samples as described above in Section 6.1. Two rounds of long PCR was performed on the DNA primers 5L (SEQ ID NO:11) and LTR3 (SEQ ID NO:19). The primer 5L contains nucleotides 7370-7395 of the MMTV BR6 genome (5'-3': CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3 contains nucleotides

BNSDOCID: <WO 9717470A1>

9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Long PCR was performed using protocols described by the manufacturer (Perkin Elmer, Foster City, CA). The amplified retroviral fragment isolated from the breast cancer sample was cloned into the TA cloning vector (Invitrogen) and automated sequencing was performed as described in Section 6.1.

7.2 Results

An approximately 2.6 kb retroviral fragment containing 1,228 bp of the 3' LTR sequence and 1,336 bp of the env gene sequence of a potential provirus was detected in a human breast carcinoma tissue sample by the long PCR technique using the 5L and LTR3 primers. The sequence of this retroviral fragment is shown in Fig. 9. (SEQ ID NO:20).

When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively, over the MMTV genomic fragment from nucleotides 7370-9937. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

8. Discussion

Search for virus-related sequences in human breast cancer has been hampered by great variation reported in previous studies, by the presence of endogenous retroviral sequences in human DNA and by the lack of sensitivity of the methods employed. The studies reported herein circumvent these deficiencies by focusing on sequences with low homology to human endogenous retroviruses, by investigating a large number of tumors and several types of controls and by using the most sensitive technology presently available.

The results indicate that unique MMTV env gene sequences were present in 38.5% of the breast cancer

BNSDOCID: <WO 9717470A1>

samples analyzed and 39.7% of archival samples of breast cancer and that these sequences were absent in normal tissues including lymphocytes from patients with positive breast cancer and in cancers other than breast. Normal breast tissue and fibroadenomas had a low frequency (1.8 to 6.9%) of positive results. When cloned and sequenced, the sequences were found to be highly homologous to MMTV env gene, but not to the endogenous retroviral sequences. Furthermore, experiments in which the cloned amplified sequences were used for hybridization with DNA from breast cancer or normal tissues revealed that homologous DNA was only present in breast cancer DNA. The results also indicate that a human breast carcinoma sample contained an about 2.6 kb MMTV-like fragment comprised of 1,336 bp of the env gene and 1,228 bp of the 3′ LTR.

The detection of MMTV env gene sequences in two fibroadenomas out of 29 and in two normal breast tissue samples out of 107 samples is of uncertain significance. Although such results could potentially be artifactual, and thus may represent false positives, they may alternatively indicate the presence of histologically unrecognized cells that were or will be neoplastic.

Ninety percent (90%) of the breast cancers tested were invasive ductal carcinomas, which reflects the prevalence of this type of neoplasm. Most patients were node-positive which is probably artifactual since it was necessary that tumor size be sufficiently large to provide an aliquot for research and tumor size correlates with node positivity.

It is unlikely that differences in homology between MMTV env gene and the cloned human sequences are generated by errors committed by the Taq polymerase. It has been estimated that the rate of nucleotide misincorporation is 1×10^{-5} per cycle (Ehrlich et al, 1991, Science 252:1643-1651) and

BNSDOCID: <WO 9717470A1>

therefore, only a total of 0.32 nucleotides misincorporated should be expected in 660 bp after 50 cycles. The differences in homology between clones from different patients is likely to represent heterogeneity of the env gene.

In contrast to earlier, ambiguous data associating MMTV-like sequences with human breast cancer, we have clearly demonstrated the existence of such sequences in breast cancer cells which cannot be explained by any known human endogenous retroviral sequence. Our data do not support the results of earlier studies which indicated that, as in the mouse, MMTV-like sequences were found in lymphocytes from two patients with breast cancer (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331). The absence of MMTV env-like sequences in lymphocytes could reflect the fate of a unique lymphocyte subset over decades between initial encounter and the appearance of clinical breast cancer; alternatively, the human disease may differ from the mouse model. Results from attempts to identify unique MMTV-like pol gene sequences have shown that they cannot be distinguished from the reverse transcriptase sequences of endogenous retroviruses (Deen, K.C. and Sweet, R.W., 1986, J. Virol. 57:422-432).

The origin of the MMTV env gene-like and 3' LTR-like sequences found in tumor DNA could be the result of integrated MMTV-like sequences from a human mammary tumor virus. Polymorphism of endogenous retroviral sequences is conceivable but can be ruled out because these sequences were not detected in lymphocytes from the positive patients, in sections of the cancerous breast from which abnormal cells were absent, or in normal breast tissue from patients with MMTV env-like positive tumors. Recombination during tumorigenesis between endogenous sequences to resemble the MMTV env genes seems highly unlikely since no known gene or viral sequence is more than 18% homologous to the

BNSDOCID: <WO 9717470A1>

660 bp sequence. The longer about 2.6 kb MMTV-like fragment detected in a human breast carcinoma had minimal homology (58% in 36 bp nd 71% in 74 bp) to endogenous human retroviral sequences. Thus, the most conservative interpretation is that our findings represent exogenous sequences from an agent similar to MMTV. Recombination between endogenous and exogenous env gene sequences are known to accelerate the development of malignancies in mice (DiFronzo, N.L. and Holland, C.A., 1993, J. Virol. 67:3763-3770). Whether the MMTV-like sequences belong to an entire acquired provirus or to an exogenous fragment integrated into endogenous sequences, is presently not known. Experiments are in progress to distinguish between these possibilities.

Several genetic alterations have been identified in human breast cancer that can be useful as markers for prevention, detection or prognosis (reviewed in Runnenbaum, I. et al., 1991, Proc. Natl. Acad. Sci. USA 88:10657-10661). The BRCA1 and BRCA2 genes have recently been described. They account for at least 5% of breast cancer and are related to familial breast cancer (Miki, Y. et al., 1994, Science 266:66-71; Wooster, R. et al., 1994, Science 265:2088-2090). We have primary evidence that familial clustering of the MMTV env gene-like sequences occurs, accounting for an even higher percentage of cancers in affected families (Holland et al. 1994, Proc. Am. Assoc. Cancer Res 35:218). The presence of MMTV-like sequences may be correlated with special clinical disease status, may provide another potential molecular marker, and may distinguish a subset of human breast cancer for which viral etiology is tenable. This has implications for epidemiology, therapy and prevention.

Various publications are cited herein, the contents of which are hereby incorporated by reference in their entireties.

BNSDOCID: <WO 9717470A1>

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: HOLLAND, JAMES

(ii) TITLE OF THE INVENTION: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(iii) NUMBER OF SEQUENCES: 20

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
(B) STREET: 30 Rockefeller Plaza
(C) CITY: New York
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10112-0228

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: NOT YET ASSIGNED
(B) FILING DATE: 08-NOV-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER 08/555,394
(B) FILING DATE: 09-NOV-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Kole, Lisa B
(B) REGISTRATION NUMBER: 35,225
(C) REFERENCE/DOCKET NUMBER: 30363-PCT - 165/

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 212-408-2628
(B) TELEFAX: 212-765-2519
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCACTGCC AGATC

BNSDOCID: <WO 9717470A1>

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGAATTCCT CACTGCCAGA TC 22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTCACTGCC AGATCGCCT 19

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TACATCTGCC TGTGTTAC 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCTACATCTG CCTGTGTTAC 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCCATACG TGCTG 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTGTGGCA TACCT 15

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGAATTCAT CTGTGGCATA CCT 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCTGTGGCA TACCTAAAGG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATCGCTTG GCTCG 15

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCAGATCGCC TTTAAGAAGG 20

BNSDOCID: <WO 9717470A1>

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAGGTAGC AGCACGTATG 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Lys Arg Pro Gly Phe Gln Glu His Glu Met Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Leu Pro His Leu Ile Asp Ile Glu Lys Arg Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

BNSDOCID: <WO 9717470A1>

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr Asn Cys Leu Asp Ser Ser Ala Tyr Asp Thr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Asp Ile Gly Asp Glu Pro Trp Phe Asp Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCCTCACTGC CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGCCTTGCG CCTTCCCTGA CCAAGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCA GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCTCAA AAGGCAGGAT TTCAGGAACA TGAGATGATT CCTACATCTC 420
TGTGTTACTT ACCCTTATGT CATATTATTA GGATTACCTC AGCTAATAGA TATAGAGAAA 480
GAGGATCTAC TTTTCATATT TCCTGTTCTT CTTGTAGATT GACTAATTGT TTAGATTCTT 540
CTGCCTACGA CTATGCAGCG ATCATAGTCA AGAGGCCGCC ATACGTGCTG CTACCTGTAG 600
ATATTGGTGA TGAACCATGG TTTGATGATT CTGCCATTCA AACCTTTAGG TATGCCACAG 660
AT 662

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid

BNSDOCID: <WO 9717470A1>

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCTCACTGN CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGACTTGCG CCTTCCCTGA CCAGGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCT GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCACAA AAGGCAGGAT TTCAAGAACA TGACATGAAT CCCTACATCT 420
CTGTGTTACT TACCCTTATG CCANANTATT AGGATTACCT CAGCTAATAG ATATAGAGGA 480
AGAGGATCTA CTTTTCATAT TTCCTGTTCT TCTTGTAGAT TGACTAATTG TTTAGATTCT 540
TCTGCCTACG ACTATGCAGC GATCATAGTC AAGAGGCCGC CATACGTGCT GCTACCTGTA 600
GATATTGGTG ATGAACCATG GTTTGATGAN NCTGCCANTC AAACCTTTAG GTATNCCACA 660
GAT 663

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGAACAGACA CAAACACACG 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

BNSDOCID: <WO 9717470A1>

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGAACAGACA CAAACACACG AGAGGTGAAT GTTAGGACTG TTGCAAGTTT ACTCAAAAAA 60
CAGCACTCTT TTATATCATG GTTTACATAA GCATTTACAT AAGACTTGGA TAAGTTCCAA 120
AAGAACATAG GAGAATAGAA CACTCAGAGC TTAGATCAAA ACATTTGATA CCAAACCAAG 180
TCAGGAAACC ACTTGTCTCA CATCCTTGTT TTAAGAACAG TTTGTGACCC TGAACTTACT 240
TAAACCTTGG GAACCGCAAN GTTGGGCTCA TAAAGGTTAT CCATTATAGC TCATGCCAAA 300
ATTATCTGCA GAAATGTGTT CCTAATTGTC TAGCCACTGC CCCCTCCCTT GGTATAATGA 360
AAATCTTTCC CCCAACGTTC ATCCCACTCC CCTAGATAAA TATAATCATG TACCTGTTGT 420
TTTATGTCGT CTTTTTCTTC CTGAGTTAAC ACACACCAAG GAGGTCTAGC TCTGGCGAGT 480
CTTTCACGAA AGGGGAGGGA TCTGTACAAC ACTTTATAGC CGTTGACTGT GACCCACCTA 540
TCGAAATTTA AATCGTATCT TCCTGTATAT GGTAGCGGGG CGTCTGTTGG TCTGTAGATG 600
TAAGTCCCGG TTGCCACCAC CTGTCTCCTA TTTTGACAAG CGTACTCCTC TTTCCCCTTT 660
TTACTTCTAG GCCTGAGGCC CTTAGTCCTT GCACCTGTTC TTCAACTGAG GTTGAGCGTC 720
TCTTTCTATT TTCTATTCCC ATTTCTAACC TTTGAATTTG AGTAAATATA GTGCTAAAAG 780
ACAAAGATTC ATTTCTTAAC ATCATGATTA ATAATCGACC TATTGGATTG GTCTTATTGG 840
TAAAAATATA ATTTTTAGCA AGCATTCTTA TTTCTATTTC TGAAGGACAA AGTCGGTGTG 900
GCTTGTAANA GGAANTTGGC TGTGGTCCTT GCCCCACGAG GAAGGTCGAG TTCTCCGAAT 960
TGTTTAGATT GTAATCTTGC ACAGAAGAGT TATTAAAAGA ATCAAGGGTG AGAGCCCTGC 1020
GAGCACGAAC CGCAACTTCC CCCAATAGCC CCAGGCAAAG CAGAGCTATG CCAAGTTTGC 1080
AGCAGANAAT GAGTATGTCT TTGTCTGATG GGCTCATCCG CGTGCACGCA GACGGGTCGT 1140
CCTTGGTGGG AAACAACCCC TTGGCTGCTT CTCTCCTAAG TGTAGGACAC TCTCGGGAGT 1200
TCAACCATTT CTGCTGCAGG CGCGGCATTT CCCCCTTTTT TCTTTTTTAA AAGAAGCACG 1260
TTAAGATCTG ACTGCACTTG GTCAAGGCTC TTCGCAAAGC ACTGGAAAAT AACGGGGAAA 1320
ATCATAAGTA CTATGACCAA AAGCAGGGCT CCAACTCCTA TAAAAATGAA ATATTGTGTT 1380
CTAATCCAAT GGATTTAAAG CCTTTACTCC ATTGGCNAAG GANTGANCCA ACCCCTGAGG 1440
TCCCTGCGTT CAAATTTTTT TGCTCNTATC CTAATCCAAT TGGTAACCCC GTTTNTTTTT 1500
GAAACTCATG TCTTCAAATG CCCAATAAAT GAGCCCTGGT TCTTTCCCAG CTCTCAGAAG 1560
CATTATACGG NANAGGTGTG ACACAGCATA AAATCATAAT TTGCATGACA CCTAGTGGAC 1620
ATTCTGGTCT TTAAGTTTGC CACATCTTGT CCCAACTCTA AAACTACTTC TTCTAAAGCA 1680
TTAAGTCTAG CTTTCAATTT TAAGTCTATT ATTCTTTGTT CAGATNAGGC TAATGTAACA 1740
TTTCTATGAA GATTATTAAC AAACGTAGCA GTTTGCATCT CCTTAACTAA GGCAGTAGTA 1800
GCTACAGCAA AGGAAGTGAT AATAGCAATT AAAGCAGATA TGCCCAGAAT AATGGCAGCG 1860
ACGAATCGCT TAGCTCGAAT TAAATCTGTG GCATACCTAA AGGTTTGAAT GGCAGAATCA 1920
TCAAACCATG GTTCATCACC AATATCTACA GGTTACAACA CATATGGCGG CCCCTTGAAT 1980
ATGAATCGCT GCATATCCGT NGGCAAAAAA TCTAACCATT ATTCCTCCTN CCNAAAAACG 2040
GGATTTGAAA NTTATNCCCC TTNCCCCNAA CCCANACCGA GGTACCCCAT AATGNGGGGG 2100
GTATCTANAA NAGGGCATAG GGGTAAGAAA AACGGCAGAG NGGGATCNTT TATGTTCNGG 2160
AAATTCNGGG TTTGGGAGAA TAAGATTCTG GAGGCTGCAA ATTAAGGGAA ACATTNTGTA 2220
TGGGGAATAG AGCAGTAAAA TCTCTATCAT GGGGATCTTT AGGGAGAATT TTCCCAGGAA 2280
CCAAGTAGGT TCNAACCCAT CNTGCTTCAT ACCATCGATG AACNTCTTTA TTGACAGGGG 2340
GAGTATAATT TCCAAATAGA TCCTTTTTGT TTTTAATCTG ATCTGACTGA TCTACACTAG 2400
GCGGGGGAAG GGAGAAATCC CAAAGTAACC CAAGGGCCCC TTTTGGAGAA AAACTCACCC 2460
CCTGGTCAGG GAAGGCGCAA GGCAACCACC GTGGAGGAGC AGACTCGTCT CCCTCCCAGA 2520
AGGCGTCCTT CTTAAAGGCG ATCTGGAGGA GCAGACTCGT CTCCCTCCCA GAAGGCGTCC 2580
TTCTTAAAGG CGATCTGG 2598

Claims

1. A composition comprising an oligonucleotide primer which may be used to detect the presence of a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 38 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 7 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

2. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATC (SEQ ID NO:1).

3. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2).

4. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATCGCCT (SEQ ID NO:3).

5. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence TACATCTGCCTGTGTTAC (SEQ ID NO:4).

6. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTACATCTGCCTGTGTTAC (SEQ ID NO:5).

7. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCGCCATACGTGCTG (SEQ ID NO:6).

BNSDOCID: <WO 9717470A1>

8. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence ATCTGTGGCATACCT (SEQ ID NO:7).

9. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8).

10. The composition of claim 1, wherein the oligonucleotide primer comprises a sequence selected from the group consisting of ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); and TACAGGTAGCAGCACGTATG (SEQ ID NO:12).

11. An essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

12. An antibody which specifically binds to the peptide of claim 11.

13. The peptide according to claim 11 which comprises the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13).

14. An antibody which specifically binds to the peptide of claim 13.

15. The peptide according to claim 11 which comprises the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14).

BNSDOCID: <WO 9717470A1>

16. A method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of 3′ LTR of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

17. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13) is detected by the binding of an antibody specific to the peptide.

18. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14) is detected by the binding of an antibody specific to the peptide.

19. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence TNCLDSSAYDTA (SEQ ID NO:15) is detected by the binding of an antibody specific to the peptide.

20. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence DIGDEPWFDD (SEQ ID NO:16) is detected by the binding of an antibody specific to the peptide.

21. A composition comprising an oligonucleotide primer which may be used to detect the presence of a

nucleic acid molecule which (i) hybridizes to a nucleic acid comprised of a sequence selected from the group consisting of the env gene and the 3' LTR of a mouse mammary tumor virus; (ii) is present in a substantial percentage of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

22. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11).

23. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CGAACAGACACAAACACACG (SEQ ID NO:19).

BNSDOCID: <WO 9717470A1>

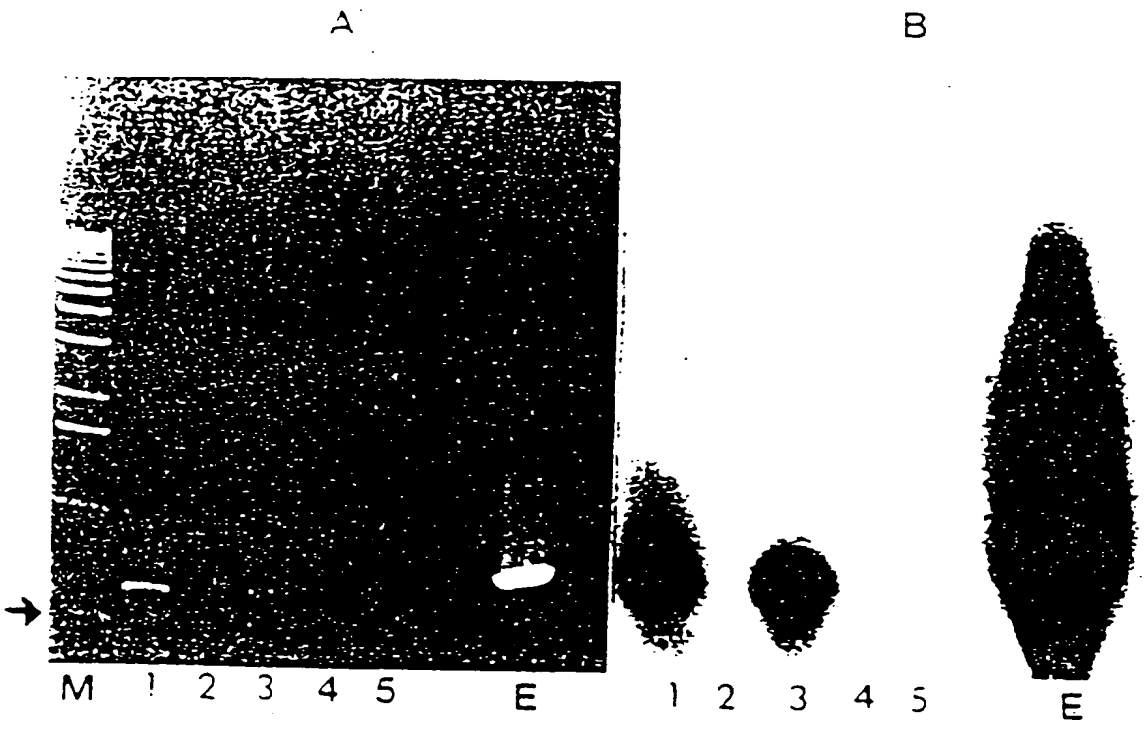
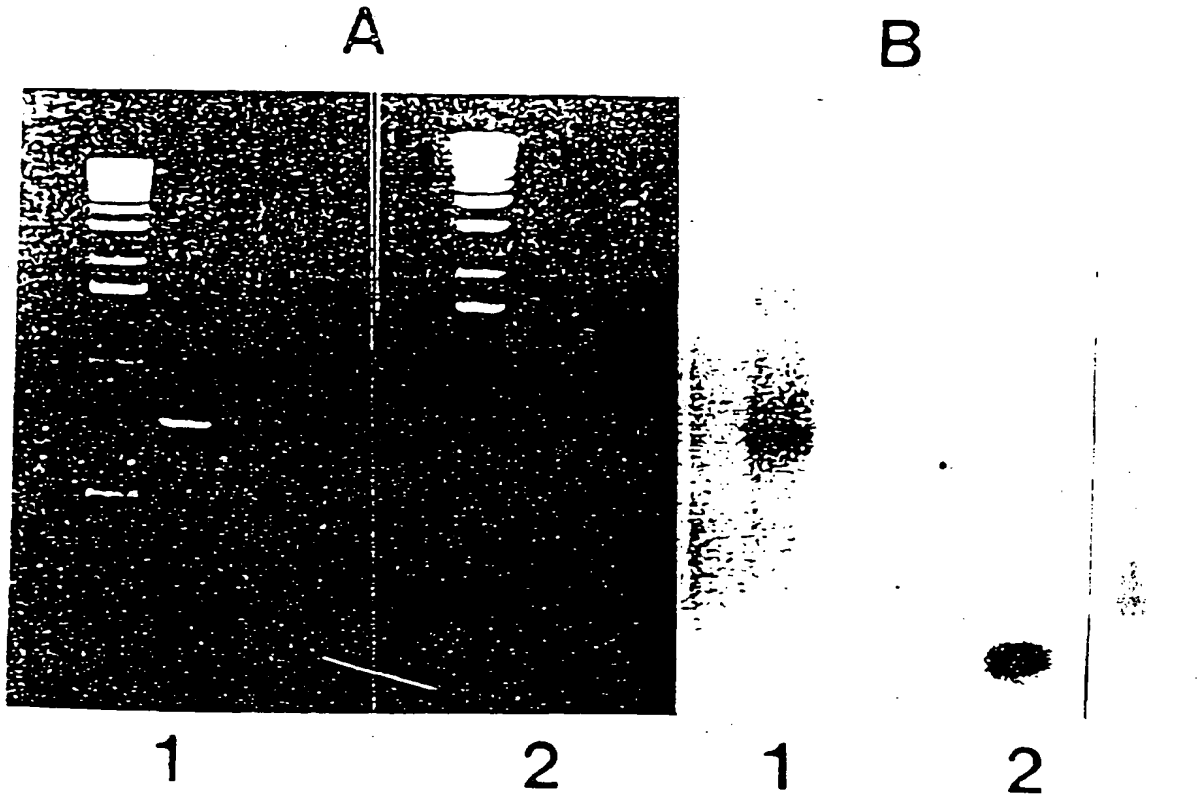


FIGURE 1

BNSDOCID: <WO 9717470A1>



2

FIGURE 2

BNSDOCID: <WO 9717470A1>

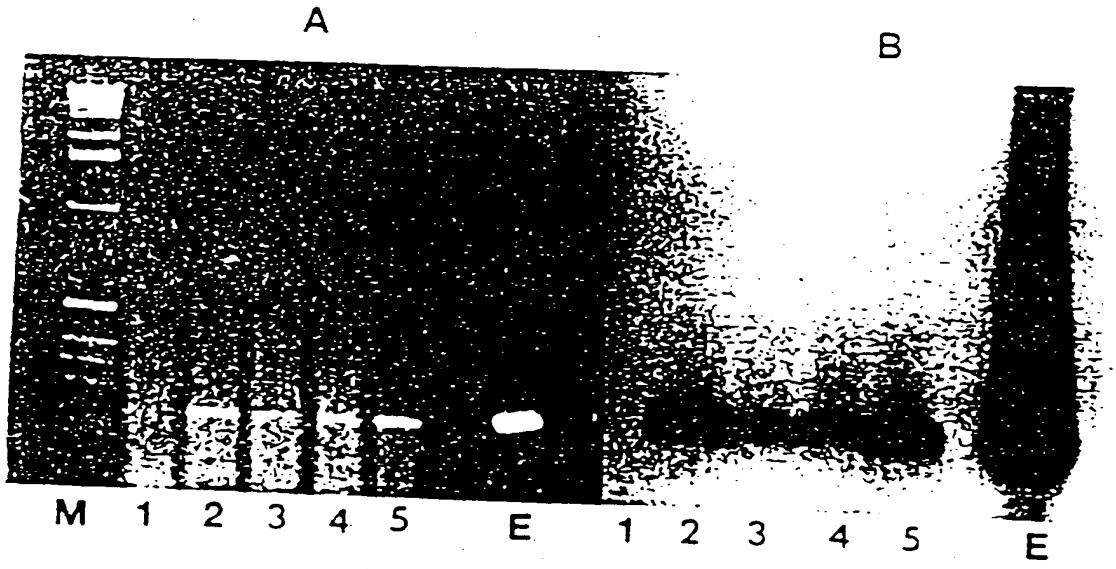


FIGURE 3

BNSDOCID: <WO 9717470A1>

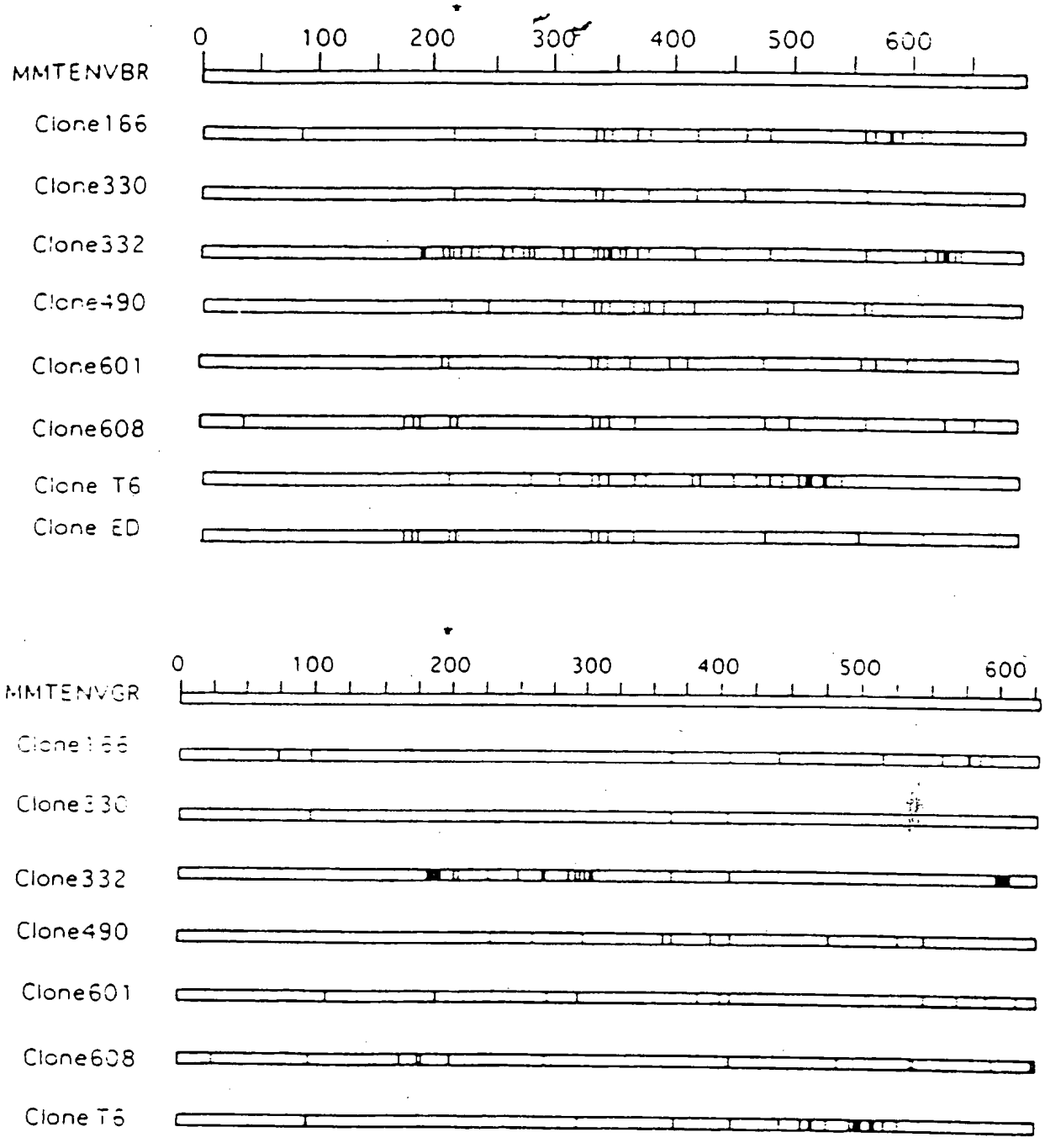


FIGURE 4

BNSDOCID: <WO 9717470A1>

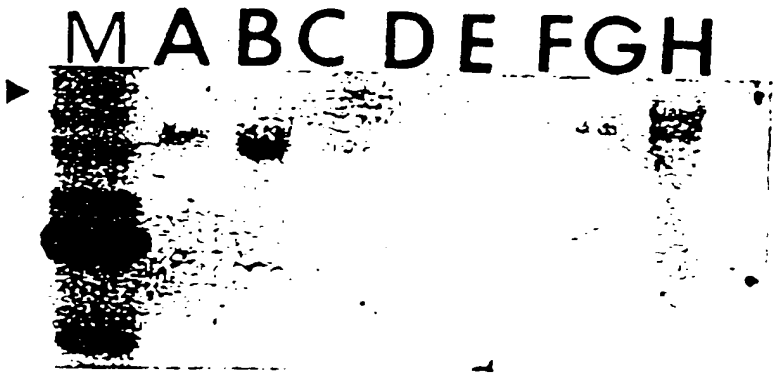
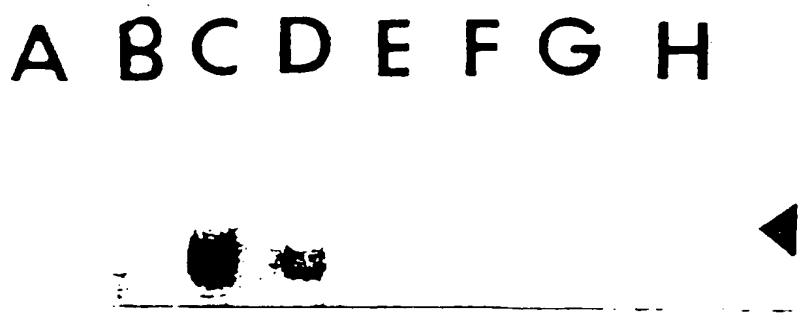


FIGURE 5

BNSDOCID: <WO 9717470A1>



6

FIGURE 6

BNSDOCID: <WO 9717470A1>

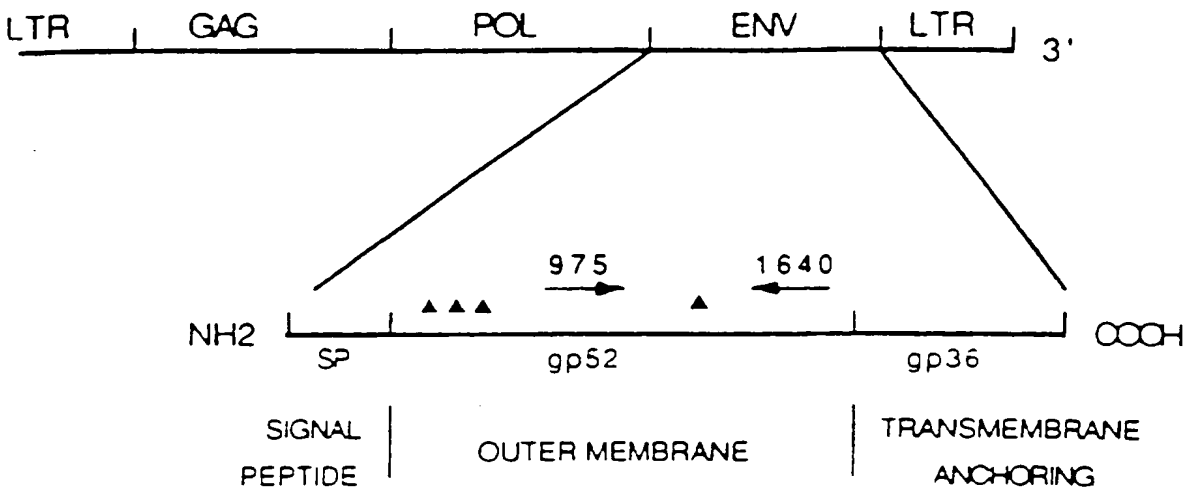


FIGURE 7

BNSDOCID: <WO 9717470A1>

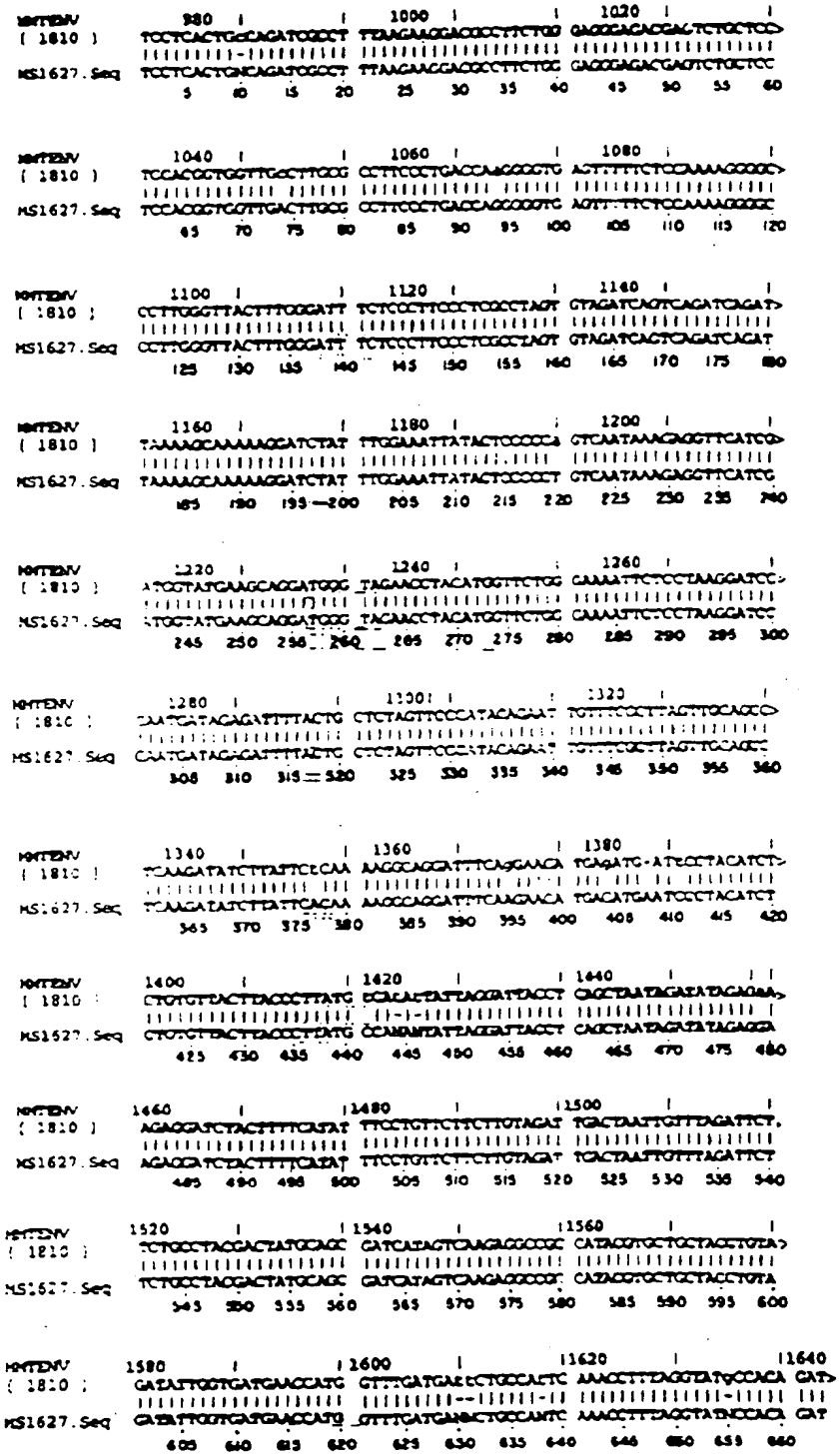


FIGURE 8

BNSDOCID: <WO 9717470A1>

CGAACAGACACNNACACACGAGAGGTGAATGTTAGGACNNTTGCAAGTTTA
CTCAAAAAACAGCACTCTTTTATATCATGGTTTACATAAGCATTTACATAAGA
CTTGGATAAGTTCCAAAAGAACATAGGAGAATAGAACACTCAGAGCTTAGAT
CAAAACATTTGATACCAAACCAAGTCAGGAAACCACTTGTCTCACATCCTTG
TTTTAAGAACAGTTTGTGACCCTGAACTTACTTAAACCTTGGGAACCGCAAN
GTTGGGCTCATAAAGGTTATCCATTATAGCTCATGCCAAAATTATCTGCAGA
AATGTGTTCCTAATTGTCTAGCCACTGCCCCCTCCCTTGGTATAATGAAAAT
CTTTCCCCCAACGTTCATCCCACTCCCCTAGATAAATATAATCATGTACCTGT
TGTTTTATGTCGTCTTTTTCTTCCTGAGTTAACACACACCAAGGAGGTCTAGC
TCTGGCGAGTCTTTCACGAAAGGGGAGGGATCTGTACAACACTTTATAGCC
GTTGACTGTGACCCACCTATCGAAATTTAAATCGTATCTTCCTGTATATGGTA
GCGGGGCGTCTGTTGGTCTGTAGATGTAAGTCCCGGTTGCCACCACCTGTC
TCCTATTTTGACAAGCGTACTCCTCTTTCCCCTTTTTACTTCTAGGCCTGAGG
CCCTTAGTCCTTGCACCTGTTCTTCAACTGAGGTTGAGCGTCTCTTTCTATTT
TCTATTCCCATTTCTAACCTTTGAATTTGAGTAAATATAGTGCTAAAAGACAA
AGATTCATTTCTTAACATCATGATTAATAATCGACCTATTGGATTGGTCTTATT
GGTAAAAATATAATTTTTAGCAAGCATTCTTATTTCTATTTCTGAAGGACAAA
GTCGGTGTGGCTTGTAANAGGAANTTGGCTGTGGTCCTTGCCCCACGAGGA
AGGTCGAGTTCTCCGAATTGTTTAGATTGTAATCTTGCACAGAAGAGTTATTA
AAAGAATCAAGGGTGAGAGCCCTGCGAGCACGAACCGCAACTTCCCCCAAT
AGCCCCAGGCAAAGCAGAGCTATGCCAAGTTTGCAGCAGANAATGAGTATG
TCTTTGTCTGATGGGCTCATCCGCGTGCACGCAGACGGGTCGTCCTTGGTG
GGAAACAACCCCTTGGCTGCTTCTCTCCTAAGTGTAGGACACTCTCGGGAG
TTCAACCATTTCTGCTGCAGGCGCGGCATTTCCCCCTTTTTTCTTTTTTAAAA
GAAGCACGTTAAGATCTGACTGCACTTGGTCAAGGCTCTTCGCAAAGCACT
GGAAAATAACGGGGAAAATCATAAGTACTATGACCAAAAGCAGGGCTCCAA
CTCCTATAAAAATGAAATATTGTGTTCTAATCCAATGGATTTAAAGCCTTTAC
TCCATTGGCNAAGGANTGANCCAACCCCTGAGGTCCCTGCGTTCAAATTTTT
TTGCTCNTATCCTAATCCAATTGGTAACCCCGTTTNTTTTTGAAACTCATGTC
TTCAAATGCCCAATAAATGAGCCCTGGTTCTTTCCCAGCTCTCAGAAGCATT
ATACGGNANAGGTGTGACACAGCATAAAATCATAATTTGCATGACACCTAGT
GGACATTCTGGTCTTTAAGTTTGCCACATCTTGTCCCAACTCTAAAACTACTT
CTTCTAAAGCATTAAGTCTAGCTTTCAATTTTAAGTCTATTATTCTTTGTTCAG
ATNAGGCTAATGTAACATTTCTATGAAGATTATTAACAAACGTAGCAGTTTGC
ATCTCCTTAACTAAGGCAGTAGTAGCTACAGCAAAGGAAGTGATAATAGCAA
TTAAAGCAGATATGCCCAGAATAATGGCAGCGACGAATCGCTTAGCTCGAAT
TAAATCTGTGGCATACCTAAAGGTTTGAATGGCAGAATCATCAAACCATGGT
TCATCACCAATATCTACAGGTTACAACACATATGGCGGCCCCTTGAATATGA
ATCGCTGCATATCCGTNGGCAAAAAATCTAACCATTATTCCTCCTNCCNAAA
AACGGGATTTGAAANTTATNCCCCTTNCCCCNAACCCANACCGAGGTACCC
CATAATGNGGGGGGTATCTANAANAGGGCATAGGGGTAAGAAAAACGGCA
GAGNGGGATCNTTTATGTTCNGGAAATTCNGGGTTTGGGAGAATAAGATTCT
GGAGGCTGCAAATTAAGGGAAACATTNTGTATGGGGAATAGAGCAGTAAAA
TCTCTATCATGGGGATCTTTAGGGAGAATTTTCCCAGGAACCAAGTAGGTTC
NAACCCATCNTGCTTCATACCATCGATGAACNTCTTTATTGACAGGGGGAGT
ATAATTTCCAAATAGATCCTTTTTGTTTTTAATCTGATCTGACTGATCTACACT
AGGCGGGGGAAGGGAGAAATCCCAAAGTAACCCAAGGGCCCCTTTTGGAG
AAAAACTCACCCCCTGGTCAGGGAAGGCGCAAGGCAACCACCGTGGAGGA
GCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGGAGG
AGCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGG

FIGURE 9

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REDMOND et al. Sequence and expression of the mouse mammary tumour virus env gene. The EMBO Journal. 1983, Volume 2, Number 1, pages 125-131. See entire document.	1-20
A	FAFF et al., Retrovirus-like particles from the human T47D cell lines are related to mouse mammary tumour virus and are of human endogenous origin. Journal of General Virology. 21 May 1992, Volume 73, pages 1087-1097. See abstract.	1-20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
04 FEBRUARY 1997	18 MAR 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DIANNE REES JAB for Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CREPIN et al. Sequences Related to Mouse Mammary Tumor Virus Genome in Tumor Cells and Lymphocytes from Patients with Breast Cancer. Biochemical and Biophysical Research Communications. 13 January 1984, Volume 118, Number 1, pages 324-331. See entire document.	1-20
A	MESA-TEJADA et al. Detection in human breast carcinomas of an antigen immunologically related to a group-specific antigen of mouse mammary tumor virus. Proceedings of the National Academy of Sciences, USA. March 1978, Volume 75, Number 3, pages 1529-1533.	1-20

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12Q 1/68, 1/70; C12P 19/34; C07H 21/02, 21/04; G01N 33/53; C07K 15/28; 5/00

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, CANCERLIT, DGENE,DRUGU, EMBASES, MEDLINE, USPATFULL, TOXLIT, TOXLINE, JAPIO, WPIDS
search terms: MMTV, mouse mammary tumor virus, PCR, hybridization, antibodies, immunoassays, Westerns, searched SEQ. ID. Nos.

BNSDOCID: <WO 9717470A1>

THIS PAGE BLANK (USPTO)

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, 1/70, C12P 19/34, C07H 21/02, 21/04, G01N 33/53, C07K 15/28, 5/00

A1

(11) International Publication Number: WO 97/17470

(43) International Publication Date: 15 May 1997 (15.05.97)

(21) International Application Number: PCT/US96/17877

(22) International Filing Date: 8 November 1996 (08.11.96)

(30) Priority Data:
08/555,394 9 November 1995 (09.11.95) US

(71)(72) Applicant and Inventor: HOLLAND, James, F. [US/US]; 31 Mamaroneck Road, Scarsdale, NY 10583 (US).

(72) Inventor: POGO, Beatriz, G., T.; 237 Nyac Avenue, Pelham, NY 10803 (US).

(74) Agents: CLARK, Richard, S. et al.; Brumbaugh, Graves, Donohue & Raymond, 30 Rockefeller Plaza, New York, NY 10112 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(57) Abstract

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

* (Referred to in PCT Gazette No. 31/1997, Section II)

BNSDOCID: <WO 9717470A1>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

BNSDOCID: <WO 9717470A1>

Description

Detection Of Mammary Tumor Virus-Like Sequences In Human Breast Cancer

Cross-Reference to Related Application

This application is a continuation-in-part application of U.S. Serial No. 08/555,394, filed November 9, 1995.

Statement Regarding Federally Sponsored Research

This invention was made with funds from the U.S. government, which has certain rights in the invention.

Introduction

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

Background of the Invention

A large body of information has accumulated about the molecular biology of MMTV (reviewed in Slagle, B.L. et al., 1987, in "Cellular and Molecular Biology of Mammary Cancer", Kidwell et al., eds., Plenum Press, NY. pp 275-306). Mouse mammary tumor virus (MMTV) is associated with a high incidence of breast cancer in certain strains of mice (over 90% among females), and has been regarded as a potential model for human disease.

The MMTV virus does not carry a transforming oncogene, but rather acts as an insertional mutagen with several proviral insertion loci designated int-1

BNSDOCID: <WO 9717470A1>

or wnt-1 (Nusse R. et al., 1982, Cell 31:99-109) int-2 (Peters, G. et al., 1983, Cell 33:369-377) int-3 (Gallahan, D. et al., 1987, J. Virol. 61:218-220) int-4 (Roelink, H. et al., 1990, Proc. Natl. acad. Sci. USA 87:4519-4523) and int-5 (Morris, V.L., et al. 1991, Oncogene Research 6:53-63), which encode for growth factors or other related proteins. These genes are not expressed in normal mammary tissue but become activated after integration of MMTV provirus into the adjacent chromosomal DNA.

The human homolog of the int-2 locus has been located on chromosome 11 (Casey, G. et al., 1986, Mol. Cell Biol. 6:502-510) and has been found amplified (in 15% of the breast cancers) and also expressed (Lidereau, R. et al., 1988, Oncogene Res 2:285-291; Zhou, D.J. et al., 1988, Oncogene 2:279-282; Liscia, D.S. et al., 1989, Oncogene 4:1219-1224; Meyers, S.L. et al., 1990, Cancer Res 50:5911-5918). It may be significant that in tumors from Parsi women, who have a high incidence of breast tumors, the int-2 locus is amplified in 50% of the cases (Barnabas-Sohi, N. et al., 1993, Breast Dis. 6:13-26). The amplification of int-2 and other genes in 11q13 is indicative of poor prognosis (Schuwring, E. et al., 1992, Cancer Research 52:5229-5234; Champeme, M-H, et al., 1995, Genes, Chromosomes and Cancer 12:128-133). Both mouse and human int-2 have been sequenced (Moore, R. et al., 1986, EMBO J 5:919-924). The gene encodes a protein of about 27 kilodaltons (KD) which shows homology to both basic and acidic fibroblast growth factors (Dickson, C. et al. 1987, Nature (London) 326:833).

However, efforts to demonstrate the presence of viruses in human breast cancer through search for viral particles, immunological cross-reactivity, or sequence homology have yielded contradictory results. Detectable MMTV env gene-related antigenic reactivity has been found in tissue sections of breast cancer

BNSDOCID: <WO 9717470A1>

(Mesa-Tejada et al., 1978, Proc. Natl. Acad. Sci. USA 75:1529-1533; Levine, P. et al., 1980, Proc. Am. Assoc. Cancer Res. 21:170; Lloyd, R. et al., 1983, Cancer 51:654-661), breast cancer cells in culture (Litvinov, S.V. and Golovkina, T.V., 1989, Acta Virologica 33:137-142), human milk (Zotter S. et al., 1980, Eur. J. Cancer 16:455-467) in sera of patients (Day, N.K. et al., 1981, Proc. Natl. Acad. Sci. USA 78:2483-2487), in cyst fluid (Witkin, S.S. et al., 1981, J. Clin. Invest. 67:216-222) and in particles produced by a human breast carcinoma cell line (Keydar, I. et al., 1984, Proc. Natl. Acad. Sci. USA 81:4188-4192). Sequence homology to MMTV has been found in human DNA under low stringency conditions of hybridization (Callahan, R. et al., 1982, Proc. Natl. Acad. Sci. USA 79:5503-5507) and RNA related to MMTV has been detected in human breast cancer cells (Axel, R. et al., 1972, Nature 235:32-36). The presence of MMTV related sequences in lymphocytes from patients with breast cancer has been reported (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331), as well as detection of reverse transcriptase (RT) activity in their monocytes (Al-Sumidaie, A.M. et al., 1988, Lancet 1:5-8). May and Westley (May and Westley, 1989, Cancer Research 49:3879-3883) have reported the presence of MMTV-like sequences arranged as tandem repeats only in DNA from breast cancer cells.

These results have been difficult to interpret, and theories linking MMTV or a related virus with human breast cancer have fallen out of favor, in view of the relatively recent discovery of human endogenous retroviral sequences ("HERs"; Westley, B. et al., 1986, J. Virol. 60:743-749; Ono, M. et al., 1986, J. Virol. 60:589-598; Faff, O. et al., 1992, J. Gen. Virology 73:1087-1097). Data which could be interpreted to demonstrate the presence of MMTV-related sequences could be more readily explained by endogenous human

BNSDOCID: <WO 9717470A1>

retroviral sequences. Adding further confusion to the picture, env-gene related antigenicity has been detected in epitopes of human proteins (Hareuveni, M. et al., 1990, Int. J. Cancer 46:1134-1135).

Brief Summary of the Invention

The present invention relates to methods for diagnosing breast cancer in humans in which the presence of mouse mammary tumor virus env gene-like sequences bears a positive correlation to the existence of malignant breast disease. It is based, at least in part, on the discovery that 38 to 40 percent of human breast cancer tissue samples tested contained gene sequences homologous to the mouse mammary tumor virus env gene that are substantially absent from other human tumors and tissues. The invention also relates to methods for diagnosing breast caner in humans in which the presence of retrovirus proviral fragments substantially homologous to the env gene and/or 3' LTR sequence of MMTV are detected. The molecular probes used in these experiments were designed to avoid cross-hybridization with endogenous human retroviral sequences. The present invention further provides for compositions of molecular probes which may be utilized in such diagnostic methods.

Brief Description of the Figures

FIGURE 1: Amplification of 660 bp of MMTV-like env gene. DNA was extracted from frozen tissues. PCR was performed using primers 1 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'^{32}P-end-labeled probe 2. Lanes 1 and 3: breast cancer; lanes 2 and 4: normal breast; lane 5: control reaction (no DNA); lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 510 bp band.

FIGURE 2: Nested PCR. A: 2% agarose gel electrophoresis. 1: Amplification of 686 bp of MMTV-like env

BNSDOCID: <WO 9717470A1>

gene sequences using primers 1 and 4 and the product of reaction A 1 as template. 2: Amplification of 250 bp of MMTV-like env gene sequences using primers 2 and 3. B, 1 and 2: Southern blot hybridization of the amplified products using probe 5'-^{32}P end-labeled probe 2a.

FIGURE 3: Amplification of 250 bp of MMTV-like env gene. DNA was extracted from paraffin-embedded tissue sections. PCR was performed using primers 2 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'-^{32}P-labeled probe 2a. Lane 1: normal breast; lanes 2 to 5: breast cancer; lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 298 bp band.

FIGURE 4: Nucleotide sequence of the cloned MMTV env gene-like sequences as compared to the env sequences of the GR and BR6 strains of MMTV using the GCG program. *:potential glycosylation site, |:mismatch to MMTV.

FIGURE 5: Southern blot hybridization of genomic DNA. DNA was extracted from frozen tissues or cell lines, digested with EcoR1 and transferred to nitrocellulose paper. Hybridization with ^{32}P-labeled clone 166. DNA from A, B, and G: env gene positive breast cancer; C and D: env negative breast cancer; E and F: normal breast; H:MCF-7 cells. M: molecular weight marker, Arrow indicates 9kb band.

FIGURE 6: Southern blot hybridization of genomic DNA. Experimental conditions as in Fig. 5. DNA from A and B: env negative breast cancer; C and D: env positive breast cancer; E: molecular weight marker (non-labelled); F. to H: normal breast. Arrow indicates position of 9 kb marker.

FIGURE 7: Map of MMTV.

FIGURE 8: Comparison of the nucleic acid sequence of mouse mammary tumor env gene ("MMTENV"), showing residues 976-1640, with the nucleic acid sequence of a

BNSDOCID: <WO 9717470A1>

representative 660 bp sequence obtained by PCR reaction of DNA from human breast cancer tissue ("MS1627").

FIGURE 9: Sequence of an about 2.6 kb MMTV-like fragment detected in a human breast carcinoma.

Detailed Description of the Invention

The present invention relates to methods and compositions for diagnosing breast cancer in humans.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which (i) hybridizes to a gene of mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. A "gene of mouse mammary tumor virus" includes, but is not limited to, the gag, pol, and env genes and the 5′ LTR and 3′ LTR sequences of MMTV. In preferred embodiments of the invention, the mouse mammary tumor virus (hereafter "MMTV") gene is the env gene and/or the 3′ LTR sequence. The term "hybridize" is used to refer to routine DNA-DNA or DNA-RNA hybridization techniques under what would be regarded, by the skilled artisan, as stringent hybridization conditions. The phrase "is present" indicates that a native form of the molecule, in an unpurified state (for example, as part of chromosomal DNA), may be detected by a standard laboratory technique, such as Southern blot or polymerase chain reaction (PCR). To be "present", the molecule may be detectable by one technique but not others. To be present in "less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects", all non-breast cancer tissue samples are considered together, but the total number of samples must be large enough to give the 5 percent

BNSDOCID: <WO 9717470A1>

value statistical significance that would be reasonable to the skilled artisan.

In order to identify such a nucleic acid molecule, the sequence of MMTV may be compared, using a computer database, to known human DNA sequences, and portions of MMTV which are less than or equal to 25 percent homologous to a human sequence may be selected for further study. The term "homologous", as used herein, refers to the presence of identical residues; for example, a first sequence is considered 25 percent homologous to a second sequence if it shares 25 percent of the residues of the first sequence. Since there is relatively greater likelihood that MMTV may bear similarity to human retroviral-like sequences, it may be preferable to evaluate whether a particular MMTV nucleic acid sequence is homologous to such sequences, for example, as endogenous human retrovirus sequences. A prototype of such viruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598).

Once an MMTV gene sequence which is less than or equal to 25 percent homologous to a human DNA sequence, such as a human endogenous retroviral sequence, is identified, the presence of nucleic acid molecules having the MMTV gene sequence in human breast cancer tissues and other tissues may be evaluated. Such evaluations may be performed either by Southern blot techniques, or, preferably, by polymerase chain reaction (PCR) techniques, which are more sensitive. In such a way, MMTV gene sequences which (i) hybridize to at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects and (ii) hybridize to less than 5 percent of DNA samples prepared from human tissues other than breast cancer tissues may be identified. A nucleic acid molecule having a MMTV gene sequence which satisfies these requirements may then be used in diagnostic methods which detect the presence of such sequence in human

BNSDOCID: <WO 9717470A1>

breast tissue by standard techniques, including PCR techniques which assay for the presence of the molecule, but also, where appropriate, Southern blot, Northern blot, or Western blot techniques, to name but a few.

In preferred embodiments, the present invention relates to a portion of MMTV localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; see Fig. 7). This about 660 bp sequence (hereafter, "the 660 bp sequence") has been found to exhibit low (16 percent) homology to the prototype human endogenous retrovirus HERV-K10, using the IBI/Pustell Sequence Analysis Program, and has also been shown to be present in 121 (38.5%) of 314 unselected breast cancer tissue samples, in cultured breast cancer cells, in 2 of 29 breast fibroadenomas (6.9%) and in 2 of 107 breast specimens from reduction mammoplasties (1.8%). The sequence was not found in normal tissues including breast, lymphocytes from breast cancer patients nor in other human cancers or cell lines (see example section, infra). Similarly, an about 250 bp sequence (hereafter "the 250 bp sequence"), between positions 1388 and 1640 in the env gene, and therefore falling within the 660 bp sequence, was detected in 60 (39.7%) of 151 breast cancer, and in one of 27 normal breast samples assayed from paraffin-embedded sections. Cloning and sequencing of the 660 bp and 250 bp sequences demonstrated that they are 95-99% homologous to MMTV env gene, but not to the known human endogenous retroviruses ("HERs") nor to other viral or human genes (<18%).

In another preferred embodiment, the present invention relates to a a nucleic acid molecule which corresponds to a retroviral genomic fragment which has substantial homology to 3′ LTR and/or env gene of the MMTV genome, and is found in a substantial percentage of breast cancer samples. By substantial percentage is

BNSDOCID: <WO 9717470A1>

meant at least 20% of tested breast cancer samples. Such a sequence is preferably comprised of the 3' LTR region and all or part of the env gene, although it may include more sequences of a retroviral genome. Most preferably, the sequence is at least comprised of an about 2.6 kb fragment which comprises the 1,228 base pair (bp) sequence of the 3' LTR sequence and 1,336 bp of the env gene sequence of MMTV (Fig. 9) (SEQ ID NO:20). When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

Retrovirus proviral sequences can be detected by PCR technology using primers derived from the MMTV genome. Such primers include primer 5L, containing the nucleotides 7376-7395 of the MMTV BR6 genome (5'-3': CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3, containing nucleotides 9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Other primers which correspond to or are homologous to MMTV sequences can be used as primers. Nucleotide fragments which correspond to or are homologous to the retroviral sequences isolated from the breast cancer samples can also be used to amplify additional retroviral fragments from the samples. Long PCR techniques can be used to amplify longer stretches of a proviral sequence.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which hybridizes to the about 2.6 kb retroviral fragment shown in Fig. 9 under stringent conditions or is at least 90 percent homologous to said fragment using the MacVector homology determining program which may be used to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods.

BNSDOCID: <WO 9717470A1>

Nucleic acids having the 660 bp sequence, the 250 bp sequence, or all or part of the about 2.6 kb sequence, may therefore be used, according to the invention, to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods. Where PCR methods are used, primers such as those listed in Table 1, below, may be utilized.

The present invention provides for compositions comprising essentially purified and isolated nucleic acid having the 660 bp sequence or the 250 bp sequence or an at least five bp, and preferably greater than or equal to ten bp, subsequence thereof. In order to maintain the desired specificity, such nucleic acid molecules may preferably contain sequence falling within the 660 bp sequence, but preferably do not contain sequences from other portions of the MMTV genome, which may, undesirably, hybridize to human sequences which are not breast cancer specific, such as HERs. Accordingly, the present invention provides for compositions wherein the isolated and purified nucleic acid molecule comprises at least a portion having a nucleic acid sequence which hybridizes to a region of the mouse mammary tumor virus env gene between residues 976 and 1640, or between residues 1388 and 1640, and wherein the isolated and purified nucleic acid molecule does not hybridize to any other region of the MMTV genome.

The 660 bp sequence, in various embodiments, may have a number of nucleotide sequences. For example, in one embodiment, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MMTENV-like sequence" (SEQ ID NO:17), which depicts the MMTV env sequence between residues 976 and 1640. In a second series of embodiments, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MS1627" (SEQ ID NO:18), which depicts a predominant sequence for the 660 bp sequence as it has been defined by

BNSDOCID: <WO 9717470A1>

sequencing analysis of the products of PCR reactions using DNA from human breast cancer tissues. In still further embodiments, the 660 bp sequence may have various other nucleotide sequences obtained by sequencing the results of PCR reactions to detect the presence of 660 bp sequence in human breast cancer tissues.

In related embodiments, the present invention provides for compositions comprising PCR primers that may be used to detect the presence of the forementioned molecules or other MMTV-like sequences. For example, the compositions may comprise one or more of the following primer molecules (5′ - 3′): CCTCACTGCCAGATC (SEQ ID NO:1); GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2); CCTCACTGCCAGATCGCCT (SEQ ID NO:3); TACATCTGCCTGTGTTAC (SEQ ID NO:4); CCTACATCTGCCTGTGTTAC (SEQ ID NO:5); CCGCCATACGTGCTG (SEQ ID NO:6); ATCTGTGGCATACCT (SEQ ID NO:7); GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8); ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); TACAGGTAGCAGCACGTATG (SEQ ID NO:12); CGAACAGACACAAACACACG (SEQ ID NO:19).

The use of such compositions and molecules in PCR and Southern blot techniques is illustrated in the non-limiting examples set forth below. The correlation between the presence of the MMTV-related nucleic acid molecules described above and breast cancer allows such molecules and compositions to be utilized in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer, wherein the detection of such nucleic acid molecules bears a positive correlation to the existence of breast cancer in a human. The results of such evaluation, together with additional clinical symptoms, signs, and laboratory test values, may be used to formulate the complete diagnosis of the patient.

BNSDOCID: <WO 9717470A1>

In further related embodiments, the present invention provides for an essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. In preferred embodiments, the MMTV gene is the env gene.

Such peptides may be used in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

The present invention also provides for antibodies (including monoclonal and polyclonal) antibodies which specifically bind to such peptides. Such antibodies may be used in methods of diagnosing breast cancer, for example, but not by way of limitation, by Western blot, immunofluorescent techniques, and so forth.

In nonlimiting embodiments of the invention, the skilled artisan may evaluate MMTV-like nucleic acid molecules for regions which would be considered likely to encode immunogenic peptides (using, for example, hydropathy plots). Such peptides may then be sequenced and used to produce antibodies that may be employed in diagnostic methods as set forth above.

For example, certain peptides encoded by portions of the 660 bp sequence have been synthesized. These

peptides, which have the sequences LKRPGFQEHEMI (SEQ ID NO:13) and GLPHLIDIEKRG (SEQ ID NO:14), have been used to produce antibodies in rabbits, and the resulting antisera have successfully identified breast cancer cells positive for MMTV env-like sequences by PCR assay. Other peptides encoded by 660 bp sequence which may be useful according to the invention include TNCLDSSAYDTA (SEQ ID NO:15) and DIGDEPWFDD (SEQ ID NO:16).

6. Example: The Detection of Mouse Mammary Tumor Virus Env Gene-Like Sequences in Human Breast Cancer Cells and Tissues

6.1. Materials and Methods

DNA from breast cancer tissue and other human cancer tissues, human placentas, normal human tissues including breast, and from several human cell lines (including eight breast cancer cell lines), and two normal breast cell lines was extracted following the procedure of Delli Bovi et al. (1986, Cancer Res. 46:6333-6338). The DNA was resuspended in a solution containing 0.05 M Tris HC1 buffer, pH 7.8, and 0.1 mM EDTA, and the amount of DNA recovered was determined by microfluorometry using Hoechst 33258 dye (Cesarone, C. et al., 1979, Anal Biochem 100:188-197). Plasmids containing the cloned genes of MMTV were obtained from the ATCC, propagated in Escherichia coli cultures and purified using anion-exchange minicolumns (Qiagen) or by precipitation with polyethylene glycol (Sambrook J., et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). Oligonucleotide primers were synthesized at the core facilities of the Brookdale Molecular Biology Center at Mount Sinai School of Medicine.

Polymerase chain reaction (PCR) was performed using Taq polymerase following the conditions recommended by the manufacturer (Perkin Elmer Cetus)

with regard to buffer, Mg^{2+} and nucleotide concentrations. Thermocycling was performed in a DNA cycler by denaturation at 94° C for 3 min. followed by either 35 or 50 cycles of 94°C for 1.5 min., 50° C for 2 min. and 72°C for 3 min. The ability of the PCR to amplify the selected regions of the MMTV env gene was tested by using as positive templates the cloned MMTV env gene and the genomic DNA of the MCF-7 cell line, since it was shown to express gp52 immunological determinants (Yang, N.S., et al., 1975, J. Natl. Cancer Inst. 61:1205-1208). Optimal Mg^{2+}, primer concentrations and requirements for the different cycling temperatures were determined with these templates. The master mix as recommended by the manufacturer was used. To detect possible contamination of the master mix components, a reaction without template was routinely tested. γ DNA and control primers provided by the manufacturer were used as control for polymerase activity. As an internal control, amplification of a 120 bp sequence estrogen receptor gene was assayed using primers designed and generously provided by Dr. Beth Schachter, (Mount Sinai School of Medicine, N.Y.). In addition, primers for actin 5 gene amplification were also used.

The product of the PCR was analyzed by electrophoresis in a 2% agarose gel. A 1 kb DNA ladder (Gibco BRL) was used to identify the size of the PCR product. To determine if the amplified sequences of the middle region of the 660 bp faithfully reproduced the sequences of the env gene of MMTV, an 18-mer sequence within the env gene was used as a probe for the 660 bp amplified sequence. The 18-mer probe was 5′ end-labeled with ^{32}P-ATP using T4 polynucleotide kinase and purified by the NENSORB nucleic acid purification cartridge (NEN). Southern blot hybridization was performed using the conditions described by (Saiki et al.,1985, Science 230:1350-1354).

BNSDOCID: <WO 9717470A1>

The product of the PCR (660 bp or 250 bp) was cloned directly from the reaction mixture into the TA cloning vector (Invitrogen) using the TA cloning kit and following the conditions recommended by the supplier. Direct cloning of the fragment isolated from the gel, was also performed. Plasmid DNA was purified by CsCl density gradient centrifugation or by precipitation with polyethylene glycol (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor), restricted with HindIII and EcoRl, electrophoresed in 2% agarose gels and transferred to nitrocellulose filters. Southern blot hybridization was carried out using a 5'-terminal labeled internal probe as described above. Cloning procedures were performed in laboratories totally separate from those where PCR was carried out. Automated DNA sequencing (using Applied Technology Sequencer Model 373A) was performed in the Brookdale Molecular Biology Center. Sequence homology was determined using the IBI MacVector GenBank and GCG Programs.

To prevent contamination of the samples, processing of human tissues was performed in a laminar flow hood. DNA extractions were done in a chemical hood located in a different room from that where PCR was performed. PCR assays were assembled in a biological hood provided with ultraviolet light. Aerosol resistant tips and dedicated positive-displacement pipettes were used throughout. All equipment used for PCR (microcentrifuge, electrophoresis apparatus, pipettors) was cleaned each time with 10% sodium hypochlorite to assure DNA decontamination (Prince and Andrus, 1992, Biotechniques 12:358-36). After the initial experiments were performed, the plasmid containing the MMTV <u>env</u> gene was frozen and never used again, to avoid contamination. However, to detect plasmid contamination from our own <u>env</u> gene clones,

primers were designed to amplify plasmid sequences. All the authentic MMTV env positive samples were then tested and found negative for plasmid contamination.

Southern blotting and hybridization were performed as described (Southern, E.M., 1975, J. Mol. Biol. 98:503-517), using the 660 bp cloned sequences labeled by the random primer procedure (Feinberg, A.P., et al., 1983, Anal. Biochem. 132:6-13). Prehybridization and hybridization were performed in a solution containing 6 x SSPE, 5% Denhardt's, 0.5% SDS, 50% formamide, 100 µg/ml denaturated salmon testis DNA, incubated for 18 hrs at 42°C, followed by washings with 2 x SSC and 0.5% SDS at room temperature and at 37°C and finally in 0.1 x SSC with 0.5% SDS at 68°C for 30 min (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). For paraffin-embedded tissue sections the conditions described by Wright and Manos (1990, in "PCR Protocols", Innis et al., eds., Academic Press, pp. 153-158) were followed using primers designed to detect a 250 bp sequence.

6.2. Results

6.2.1. Selection of Specific MMTV Env Gene Sequences

A computer search for MMTV env gene homologous sequences was first performed, since sequence homology between the human endogenous retroviral sequences and MMTV had been described. The prototype of this group of human endogenous retroviruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598). The sequences of the env gene of MMTV (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504) were aligned with sequences of the env gene of the human endogenous retrovirus HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598), using the IBI/Pustell Sequence Analysis Program. A region of 660 bp of low homology (16%) was localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504). This

BNSDOCID: <WO 9717470A1>

internal domain of the outer membrane of the env gene has only one glycosylation site and is highly conserved between strains. Two primers comprising 15 bp sequences at positions 976-990 (primer 1) and 1626-1640 (primer 3) were first synthesized. Later longer primers were synthesized (1N and 3N). An 18-mer sequence in the middle of the 660 bp MMTV env region (1388-1405) (primer 2) was used as a probe to identify the 660 bp sequence. A second oligomer probe was synthesized comprising the sequence 1554 to 1568 (primer 2a) to be used for hybridization when a sequence of around 250 bp (between positions 1388 and 1640) was amplified. For nested PCR reactions (Mullis, K.B. and Faloona, F.A., 1987, Meth Enzymol 155:335-350), another primer comprising sequences 1647 to 1661 (primer 4) was synthesized to be used with primer 1 in the first reaction and primers 2 and 3 in the second. Modified primers with GC clamps and extra sequences were also synthesized and used in the PCR (primers 1a and 3a). Another set of primers comprising sequences 974 to 1003 (5L) and 1558 to 1577 (3L) were subsequently developed because their Tm's matched and provided better amplification than the original primers. The sequences are represented in Table 1. All of them were productive in amplification reactions.

BNSDOCID: <WO 9717470A1>

Tabl 1. Primer and prob sequ nces and location in mouse mammary tumor virus env gene

Designation	Sequence (5'-3')	Location
1	CCTCACTGCCAGATC	976-990
1a	GGGAATTCCTCACTGCCAGATC	976-990
1N	CCTCACTGCCAGATCGCCT	976-993
2	TACATCTGCCTGTGTTAC	1388-1405
2N	CCTACATCTGCCTGTGTTAC	1386-1405
2a	CCGCCATACGTGCTG	1554-1568
3	ATCTGTGGCATACCT	1640-1626
3a	GGGAATTCATCTGTGGCATACCT	1640-1626
3N	ATCTGTGGCATACCTAAAGG	1640-1621
4	GAATCGCTTGGCTCG	1661-1647
5L	CCAGATCGCCTTTAAGAAGG	984-1003
3L	TACAGGTAGCAGCACGTATG	1558-1577

6.2.2. Detection of MMTV-Like Env Gene Sequences in Human Breast Tumor DNA

PCR was performed on DNA extracted from breast cancer tissues, normal breast tissues and from the plasmid containing the env gene of MMTV, using primers 1 and 3. Photographs of the ethidium bromide stained gels of the PCR product reveal the presence of an approximately 660 bp sequence in some of the tumors, (Fig. 1A, lanes 1 and 3) but not in the normal tissue samples (Fig. 1A, lanes 2 and 4). As a positive control the MMTV env gene was also amplified (Fig. 1A, lane E). Similar results were obtained with modified primers 1a, 3a, 3L and 5L. Southern blot hybridization of the gel with ^{32}P-labeled 18-mer oligonucleotide (primer 2) indicated that this internal sequence was present in the amplified material (Fig. 1B) and that the bands in the gel were not artifactual.

Our initial effort was to analyze a representative sample of breast cancer specimens as well as normal

BNSDOCID: <WO 9717470A1>

tissues and other tumors. To date 343 breast tumors have been processed, DNA extracted and PCR preformed. Of these 343 tumors, 314 were carcinomas and 29 were fibroadenomas. Amplification of sequences of 660 bp was observed in 121 of the carcinomas (38.5%) and in 2 of the 29 fibroadenomas (6.9%). These sequences were confirmed to be MMTV env gene-like sequences by hybridization with the labeled specific probe containing the internal sequences. These sequences were not detected in the DNAs extracted from 20 normal organs, 23 cancers from other organs and 26 samples of blood lymphocytes including 7 from breast cancer patients whose breast specimens were positive. From 107 samples of normal breast obtained from reduction mammoplasties, 2 were positive (1.8%). In addition to DNA from lymphocytes from seven positive patients, DNA from their normal breast tissue of the operated breast was tested in 4 cases. All were negative (Table 2). Finally, DNA of the MCF-7, and ED (a cell line developed in our laboratory from the pleural effusion of a patient with an env -positive breast tumor) breast cancer cell lines were shown to contain the 660 bp MMTV env gene-like sequences (Table 3), while four other breast cancer cell lines were positive only for the 250 bp sequence (T47-D, BT-474, BT-20 and MDA-MB-231).

BNSDOCID: <WO 9717470A1>

Table 2. Detection of MMTV env g n -like sequences in human DNA xtract d from fresh or frozen tissues

Sample	Number	MMTV env gene sequences	% Positive
Breast Carcinomas	314	121	38.5%
Breast Fibroadenomas	29	2	6.9%
Normal Breasts	107	2	1.8%
*Normal Breasts	4	negative	
Tumors other than breast	23	negative	
Normal tissues	20	negative	
Lymphocytes	26	negative	
**Lymphocytes	7	negative	

* Histologically normal tissue from same breast as positive cancer.

** Lymphocytes from breast cancer patients who were positive for MMTV env gene sequences in the tumor.

BNSDOCID: <WO 9717470A1>

Table 3. Detection of MMTV *env* gene-like sequences in DNA from human cell lines in culture

Human Cell Lines		MMTV *env* gene sequence
MC-7	(breast carcinoma)	positive
T47-D	" "	negative
BT-20	" "	negative
MDA-MB-231	" "	negative
ZR-75-1	" "	negative
SK-BR 3	" "	negative
BT474	" "	negative
ED	" "	positive
MCF-10	(normal breast)	negative
HB-447	" "	negative
HL-60	(promyelocytic leukemia)	negative
K562	(erythroleukemia)	negative
Jurkat	(T cell leukemia)	negative
Hep 6-2	(hepatoma)	negative

The nested polymerase reaction was used in several instances to increase sensitivity and specificity, thus reducing the probability of false positives. In Fig. 2, results of a representative nested reaction are shown using primers 1 and 4 in the first reaction (Fig. 2A) and 2 and 3 for the 2nd reaction. The specificity of the reaction can be seen in the 2nd amplification (Fig. 2B).

To study a large number of samples and to be able to perform archival studies, PCR of paraffin-embedded tissue sections was also carried out. Primers 2 and 3 were used to amplify a 250 bp sequence within the 660 bp stretch when DNA was extracted from paraffin-embedded tissue sections since larger size sequences are difficult to amplify after fixation. Tumor DNA was amplified (Fig. 3A, lanes 2-5) whereas normal breast DNA was not (Fig. 3A, lane 1). The identification of

BNSDOCID: <WO 9717470A1>

this 250 bp sequence with the MMTV-like env gene was confirmed by hybridization with an internal probe (primer 2a) as shown in Fig. 3B. Using this procedure we have analyzed 151 breast cancer samples and found that 60 (39.7%) possess the 250 bp sequence. Of the 27 normal breast samples obtained from reduction mammoplasties assayed by this procedure, one was positive (3.7%). These results, in conjunction with those obtained from lymphocytes and from normal breast tissue of patients whose breast cancer was PCR positive, indicate that MMTV-like sequences are present in a significant number of human breast cancer DNA which cannot be explained by DNA polymorphism.

6.2.3. Cloning and Sequencing of the MMTV-Like Env Gene Sequences

To find out whether there was homology to MMTV env gene throughout the whole 660 bp stretch, the product of the PCR from 8 different tumors was cloned and sequenced. In Fig. 4 the sequence of different clones comprising around 600 bp are represented, as aligned to the MMTV env gene sequence of the GR and BR6 strains (Redmon, S. and Dickson, C., 1983, EMBO J. 2:125-131). This domain of the env gene in the GR strain is 100% homologous to the C_3H strain and 98% to the BR6 strain (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; Moore, R. et al., 1987, J. Virol. 61:480-490). Evaluation of the clones indicated that homology to MMTV env gene varied from 95% to 99%. Another seven clones comprising only 250 bp were also sequenced. Homology to MMTV env gene varied from 95% to 99% (data not shown). When compared to the human endogenous provirus HERV-K10, the homology of all the clones was less than 15%. When compared against all known viral and human genes (more than 130,000 entries) using the 1B1 MacVector GenBank and GCG programs, the highest homology recorded was 18%.

BNSDOCID: <WO 9717470A1>

6.2.4. Southern Blot Analysis Using Cloned Sequences

To investigate whether the env gene-like sequences were present in human DNA, Southern blot hybridization was performed using the cloned sequence as probe. DNAs from normal breast tissues, env positive or negative breast tumors, tumors other than breast and breast cancer cell lines were restricted with EcoRI and in some instances with Pstl, Bglll or Kpnl. EcoRl is a frequent cutter restriction enzyme that digests MMTV proviral DNA between env and pol genes. Four different cloned 660 bp sequences were used as probes after labeling with ^{32}P by random prime-labeling. Results of some of the Southern blot hybridization experiments are shown in Fig. 5. They reveal the presence of a labeled restriction fragment migrating at approximately 7-8 kb in breast cancer DNA, in ED and two fragments in MCF-7 cells. Different restriction patterns were observed with the other three enzymes. The 660 bp sequence was absent in 10 normal tissues, 10 fibroadenomas and 10 tumors from other tissues. It is important to emphasize that hybridization conditions for these experiments were stringent (as described in Section 6.1) to avoid interference with endogenous sequences that might interact with the probes.

7. Example: Detection of a Retrovirus Proviral Fragment in Human Breast Cancer Cells and Tissues

7.1. Materials and Methods

To detect longer retrovirus proviral fragments in breast cancer samples, DNA was extracted from breast cancer carcinoma tissue samples as described above in Section 6.1. Two rounds of long PCR was performed on the DNA primers 5L (SEQ ID NO:11) and LTR3 (SEQ ID NO:19). The primer 5L contains nucleotides 7370-7395 of the MMTV BR6 genome (5'-3': CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3 contains nucleotides

BNSDOCID: <WO 9717470A1>

9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Long PCR was performed using protocols described by the manufacturer (Perkin Elmer, Foster City, CA). The amplified retroviral fragment isolated from the breast cancer sample was cloned into the TA cloning vector (Invitrogen) and automated sequencing was performed as described in Section 6.1.

7.2 Results

An approximately 2.6 kb retroviral fragment containing 1,228 bp of the 3' LTR sequence and 1,336 bp of the env gene sequence of a potential provirus was detected in a human breast carcinoma tissue sample by the long PCR technique using the 5L and LTR3 primers. The sequence of this retroviral fragment is shown in Fig. 9. (SEQ ID NO:20).

When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively, over the MMTV genomic fragment from nucleotides 7370-9937. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

8. Discussion

Search for virus-related sequences in human breast cancer has been hampered by great variation reported in previous studies, by the presence of endogenous retroviral sequences in human DNA and by the lack of sensitivity of the methods employed. The studies reported herein circumvent these deficiencies by focusing on sequences with low homology to human endogenous retroviruses, by investigating a large number of tumors and several types of controls and by using the most sensitive technology presently available.

The results indicate that unique MMTV env g ne sequences were present in 38.5% of the breast cancer

BNSDOCID: <WO 9717470A1>

samples analyzed and 39.7% of archival samples of breast cancer and that these sequences were absent in normal tissues including lymphocytes from patients with positive breast cancer and in cancers other than breast. Normal breast tissue and fibroadenomas had a low frequency (1.8 to 6.9%) of positive results. When cloned and sequenced, the sequences were found to be highly homologous to MMTV env gene, but not to the endogenous retroviral sequences. Furthermore, experiments in which the cloned amplified sequences were used for hybridization with DNA from breast cancer or normal tissues revealed that homologous DNA was only present in breast cancer DNA. The results also indicate that a human breast carcinoma sample contained an about 2.6 kb MMTV-like fragment comprised of 1,336 bp of the env gene and 1,228 bp of the 3' LTR.

The detection of MMTV env gene sequences in two fibroadenomas out of 29 and in two normal breast tissue samples out of 107 samples is of uncertain significance. Although such results could potentially be artifactual, and thus may represent false positives, they may alternatively indicate the presence of histologically unrecognized cells that were or will be neoplastic.

Ninety percent (90%) of the breast cancers tested were invasive ductal carcinomas, which reflects the prevalence of this type of neoplasm. Most patients were node-positive which is probably artifactual since it was necessary that tumor size be sufficiently large to provide an aliquot for research and tumor size correlates with node positivity.

It is unlikely that differences in homology between MMTV env gene and the cloned human sequences are generated by errors committed by the Taq polymerase. It has been estimated that the rate of nucleotide misincorporation is 1×10^{-5} per cycle (Ehrlich et al, 1991, Science 252:1643-1651) and

BNSDOCID: <WO 9717470A1>

therefore, only a total of 0.32 nucleotides misincorporated should be expected in 660 bp after 50 cycles. The differences in homology between clones from different patients is likely to represent heterogeneity of the env gene.

In contrast to earlier, ambiguous data associating MMTV-like sequences with human breast cancer, we have clearly demonstrated the existence of such sequences in breast cancer cells which cannot be explained by any known human endogenous retroviral sequence. Our data do not support the results of earlier studies which indicated that, as in the mouse, MMTV-like sequences were found in lymphocytes from two patients with breast cancer (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331). The absence of MMTV env-like sequences in lymphocytes could reflect the fate of a unique lymphocyte subset over decades between initial encounter and the appearance of clinical breast cancer; alternatively, the human disease may differ from the mouse model. Results from attempts to identify unique MMTV-like pol gene sequences have shown that they cannot be distinguished from the reverse transcriptase sequences of endogenous retroviruses (Deen, K.C. and Sweet, R.W., 1986, J. Virol. 57:422-432).

The origin of the MMTV env gene-like and 3' LTR-like sequences found in tumor DNA could be the result of integrated MMTV-like sequences from a human mammary tumor virus. Polymorphism of endogenous retroviral sequences is conceivable but can be ruled out because these sequences were not detected in lymphocytes from the positive patients, in sections of the cancerous breast from which abnormal cells were absent, or in normal breast tissue from patients with MMTV env-like positive tumors. Recombination during tumorigenesis between endogenous sequences to resemble the MMTV env genes seems highly unlikely since no known gene or viral sequence is more than 18% homologous to the

BNSDOCID: <WO 9717470A1>

660 bp sequence. The longer about 2.6 kb MMTV-like fragment detected in a human breast carcinoma had minimal homology (58% in 36 bp nd 71% in 74 bp) to endogenous human retroviral sequences. Thus, the most conservative interpretation is that our findings represent exogenous sequences from an agent similar to MMTV. Recombination between endogenous and exogenous env gene sequences are known to accelerate the development of malignancies in mice (DiFronzo, N.L. and Holland, C.A., 1993, J. Virol. 67:3763-3770). Whether the MMTV-like sequences belong to an entire acquired provirus or to an exogenous fragment integrated into endogenous sequences, is presently not known. Experiments are in progress to distinguish between these possibilities.

Several genetic alterations have been identified in human breast cancer that can be useful as markers for prevention, detection or prognosis (reviewed in Runnenbaum, I. et al., 1991, Proc. Natl. Acad. Sci. USA 88:10657-10661). The BRCA1 and BRCA2 genes have recently been described. They account for at least 5% of breast cancer and are related to familial breast cancer (Miki, Y. et al., 1994, Science 266:66-71; Wooster, R. et al., 1994, Science 265:2088-2090). We have primary evidence that familial clustering of the MMTV env gene-like sequences occurs, accounting for an even higher percentage of cancers in affected families (Holland et al. 1994, Proc. Am. Assoc. Cancer Res 35:218). The presence of MMTV-like sequences may be correlated with special clinical disease status, may provide another potential molecular marker, and may distinguish a subset of human breast cancer for which viral etiology is tenable. This has implications for epidemiology, therapy and prevention.

Various publications are cited herein, the contents of which are hereby incorporated by reference in their entireties.

BNSDOCID: <WO 9717470A1>

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: HOLLAND, JAMES

(ii) TITLE OF THE INVENTION: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(iii) NUMBER OF SEQUENCES: 20

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
(B) STREET: 30 Rockefeller Plaza
(C) CITY: New York
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10112-0228

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: NOT YET ASSIGNED
(B) FILING DATE: 08-NOV-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER 08/555,394
(B) FILING DATE: 09-NOV-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Kole, Lisa B
(B) REGISTRATION NUMBER: 35,225
(C) REFERENCE/DOCKET NUMBER: 30363-PCT - 165/

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 212-408-2628
(B) TELEFAX: 212-765-2519
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCACTGCC AGATC 15

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGAATTCCT CACTGCCAGA TC 22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTCACTGCC AGATCGCCT 19

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TACATCTGCC TGTGTTAC 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCTACATCTG CCTGTGTTAC 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCCATACG TGCTG 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTGTGGCA TACCT 15

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

BNSDOCID: <WO 9717470A1>

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGAATTCAT CTGTGGCATA CCT 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCTGTGGCA TACCTAAAGG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATCGCTTG GCTCG 15

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCAGATCGCC TTTAAGAAGG 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAGGTAGC AGCACGTATG 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Lys Arg Pro Gly Phe Gln Glu His Glu Met Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Leu Pro His Leu Ile Asp Ile Glu Lys Arg Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: lin ar

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr Asn Cys Leu Asp Ser Ser Ala Tyr Asp Thr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Asp Ile Gly Asp Glu Pro Trp Phe Asp Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCCTCACTGC CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGCCTTGCG CCTTCCCTGA CCAAGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCA GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCTCAA AAGGCAGGAT TTCAGGAACA TGAGATGATT CCTACATCTC 420
TGTGTTACTT ACCCTTATGT CATATTATTA GGATTACCTC AGCTAATAGA TATAGAGAAA 480
GAGGATCTAC TTTTCATATT TCCTGTTCTT CTTGTAGATT GACTAATTGT TTAGATTCTT 540
CTGCCTACGA CTATGCAGCG ATCATAGTCA AGAGGCCGCC ATACGTGCTG CTACCTGTAG 600
ATATTGGTGA TGAACCATGG TTTGATGATT CTGCCATTCA AACCTTTAGG TATGCCACAG 660
AT 662

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid

BNSDOCID: <WO 9717470A1>

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCTCACTGN CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGACTTGCG CCTTCCCTGA CCAGGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCT GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCACAA AAGGCAGGAT TTCAAGAACA TGACATGAAT CCCTACATCT 420
CTGTGTTACT TACCCTTATG CCANANTATT AGGATTACCT CAGCTAATAG ATATAGAGGA 480
AGAGGATCTA CTTTTCATAT TTCCTGTTCT TCTTGTAGAT TGACTAATTG TTTAGATTCT 540
TCTGCCTACG ACTATGCAGC GATCATAGTC AAGAGGCCGC CATACGTGCT GCTACCTGTA 600
GATATTGGTG ATGAACCATG GTTTGATGAN NCTGCCANTC AAACCTTTAG GTATNCCACA 660
GAT 663

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGAACAGACA CAAACACACG 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

BNSDOCID: <WO 9717470A1>

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGAACAGACA CAAACACACG AGAGGTGAAT GTTAGGACTG TTGCAAGTTT ACTCAAAAAA 60
CAGCACTCTT TTATATCATG GTTTACATAA GCATTTACAT AAGACTTGGA TAAGTTCCAA 120
AAGAACATAG GAGAATAGAA CACTCAGAGC TTAGATCAAA ACATTTGATA CCAAACCAAG 180
TCAGGAAACC ACTTGTCTCA CATCCTTGTT TTAAGAACAG TTTGTGACCC TGAACTTACT 240
TAAACCTTGG GAACCGCAAN GTTGGGCTCA TAAAGGTTAT CCATTATAGC TCATGCCAAA 300
ATTATCTGCA GAAATGTGTT CCTAATTGTC TAGCCACTGC CCCCTCCCTT GGTATAATGA 360
AAATCTTTCC CCCAACGTTC ATCCCACTCC CCTAGATAAA TATAATCATG TACCTGTTGT 420
TTTATGTCGT CTTTTTCTTC CTGAGTTAAC ACACACCAAG GAGGTCTAGC TCTGGCGAGT 480
CTTTCACGAA AGGGGAGGGA TCTGTACAAC ACTTTATAGC CGTTGACTGT GACCCACCTA 540
TCGAAATTTA AATCGTATCT TCCTGTATAT GGTAGCGGGG CGTCTGTTGG TCTGTAGATG 600
TAAGTCCCGG TTGCCACCAC CTGTCTCCTA TTTTGACAAG CGTACTCCTC TTTCCCCTTT 660
TTACTTCTAG GCCTGAGGCC CTTAGTCCTT GCACCTGTTC TTCAACTGAG GTTGAGCGTC 720
TCTTTCTATT TTCTATTCCC ATTTCTAACC TTTGAATTTG AGTAAATATA GTGCTAAAAG 780
ACAAAGATTC ATTTCTTAAC ATCATGATTA ATAATCGACC TATTGGATTG GTCTTATTGG 840
TAAAAATATA ATTTTTAGCA AGCATTCTTA TTTCTATTTC TGAAGGACAA AGTCGGTGTG 900
GCTTGTAANA GGAANTTGGC TGTGGTCCTT GCCCCACGAG GAAGGTCGAG TTCTCCGAAT 960
TGTTTAGATT GTAATCTTGC ACAGAAGAGT TATTAAAAGA ATCAAGGGTG AGAGCCCTGC 1020
GAGCACGAAC CGCAACTTCC CCCAATAGCC CCAGGCAAAG CAGAGCTATG CCAAGTTTGC 1080
AGCAGANAAT GAGTATGTCT TTGTCTGATG GGCTCATCCG CGTGCACGCA GACGGGTCGT 1140
CCTTGGTGGG AAACAACCCC TTGGCTGCTT CTCTCCTAAG TGTAGGACAC TCTCGGGAGT 1200
TCAACCATTT CTGCTGCAGG CGCGGCATTT CCCCCTTTTT TCTTTTTTAA AAGAAGCACG 1260
TTAAGATCTG ACTGCACTTG GTCAAGGCTC TTCGCAAAGC ACTGGAAAAT AACGGGGAAA 1320
ATCATAAGTA CTATGACCAA AAGCAGGGCT CCAACTCCTA TAAAAATGAA ATATTGTGTT 1380
CTAATCCAAT GGATTTAAAG CCTTTACTCC ATTGGCNAAG GANTGANCCA ACCCCTGAGG 1440
TCCCTGCGTT CAAATTTTTT TGCTCNTATC CTAATCCAAT TGGTAACCCC GTTTNTTTTT 1500
GAAACTCATG TCTTCAAATG CCCAATAAAT GAGCCCTGGT TCTTTCCCAG CTCTCAGAAG 1560
CATTATACGG NANAGGTGTG ACACAGCATA AAATCATAAT TTGCATGACA CCTAGTGGAC 1620
ATTCTGGTCT TTAAGTTTGC CACATCTTGT CCCAACTCTA AAACTACTTC TTCTAAAGCA 1680
TTAAGTCTAG CTTTCAATTT TAAGTCTATT ATTCTTTGTT CAGATNAGGC TAATGTAACA 1740
TTTCTATGAA GATTATTAAC AAACGTAGCA GTTTGCATCT CCTTAACTAA GGCAGTAGTA 1800
GCTACAGCAA AGGAAGTGAT AATAGCAATT AAAGCAGATA TGCCCAGAAT AATGGCAGCG 1860
ACGAATCGCT TAGCTCGAAT TAAATCTGTG GCATACCTAA AGGTTTGAAT GGCAGAATCA 1920
TCAAACCATG GTTCATCACC AATATCTACA GGTTACAACA CATATGGCGG CCCCTTGAAT 1980
ATGAATCGCT GCATATCCGT NGGCAAAAAA TCTAACCATT ATTCCTCCTN CCNAAAAACG 2040
GGATTTGAAA NTTATNCCCC TTNCCCCNAA CCCANACCGA GGTACCCCAT AATGNGGGGG 2100
GTATCTANAA NAGGGCATAG GGGTAAGAAA AACGGCAGAG NGGGATCNTT TATGTTCNGG 2160
AAATTCNGGG TTTGGGAGAA TAAGATTCTG GAGGCTGCAA ATTAAGGGAA ACATTNTGTA 2220
TGGGGAATAG AGCAGTAAAA TCTCTATCAT GGGGATCTTT AGGGAGAATT TTCCCAGGAA 2280
CCAAGTAGGT TCNAACCCAT CNTGCTTCAT ACCATCGATG AACNTCTTTA TTGACAGGGG 2340
GAGTATAATT TCCAAATAGA TCCTTTTTGT TTTTAATCTG ATCTGACTGA TCTACACTAG 2400
GCGGGGGAAG GGAGAAATCC CAAAGTAACC CAAGGGCCCC TTTTGGAGAA AAACTCACCC 2460
CCTGGTCAGG GAAGGCGCAA GGCAACCACC GTGGAGGAGC AGACTCGTCT CCCTCCCAGA 2520
AGGCGTCCTT CTTAAAGGCG ATCTGGAGGA GCAGACTCGT CTCCCTCCCA GAAGGCGTCC 2580
TTCTTAAAGG CGATCTGG 2598

BNSDOCID: <WO 9717470A1>

Claims

1. A composition comprising an oligonucleotide primer which may be used to detect the presence of a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 38 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 7 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

2. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATC (SEQ ID NO:1).

3. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2).

4. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATCGCCT (SEQ ID NO:3).

5. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence TACATCTGCCTGTGTTAC (SEQ ID NO:4).

6. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTACATCTGCCTGTGTTAC (SEQ ID NO:5).

7. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCGCCATACGTGCTG (SEQ ID NO:6).

BNSDOCID: <WO 9717470A1>

8. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence ATCTGTGGCATACCT (SEQ ID NO:7).

9. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8).

10. The composition of claim 1, wherein the oligonucleotide primer comprises a sequence selected from the group consisting of ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); and TACAGGTAGCAGCACGTATG (SEQ ID NO:12).

11. An essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

12. An antibody which specifically binds to the peptide of claim 11.

13. The peptide according to claim 11 which comprises the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13).

14. An antibody which specifically binds to the peptide of claim 13.

15. The peptide according to claim 11 which comprises the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14).

BNSDOCID: <WO 9717470A1>

16. A method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of 3' LTR of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

17. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13) is detected by the binding of an antibody specific to the peptide.

18. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14) is detected by the binding of an antibody specific to the peptide.

19. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence TNCLDSSAYDTA (SEQ ID NO:15) is detected by the binding of an antibody specific to the peptide.

20. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence DIGDEPWFDD (SEQ ID NO:16) is detected by the binding of an antibody specific to the peptide.

21. A composition comprising an oligonucleotide primer which may be used to detect the presence of a

BNSDOCID: <WO 9717470A1>

nucleic acid molecule which (i) hybridizes to a nucleic acid comprised of a sequence selected from the group consisting of the env gene and the 3' LTR of a mouse mammary tumor virus; (ii) is present in a substantial percentage of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

22. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11).

23. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CGAACAGACACAAACACACG (SEQ ID NO:19).

BNSDOCID: <WO 9717470A1>

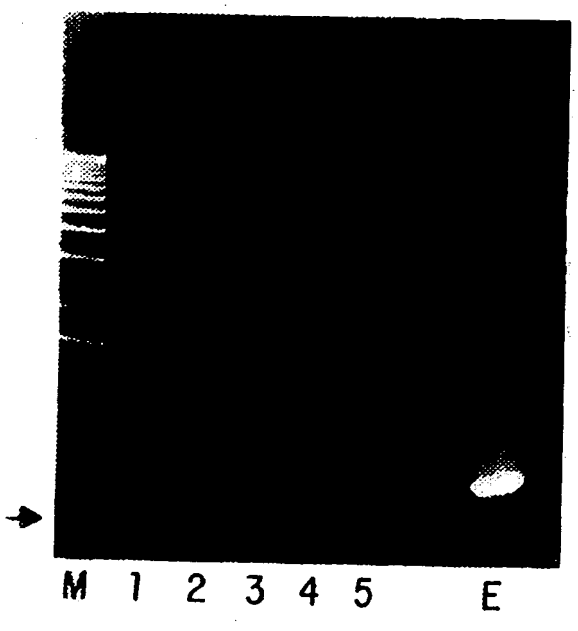


FIG. 1A

FIG. 1B

BNSDOCID: <WO 9717470A1>

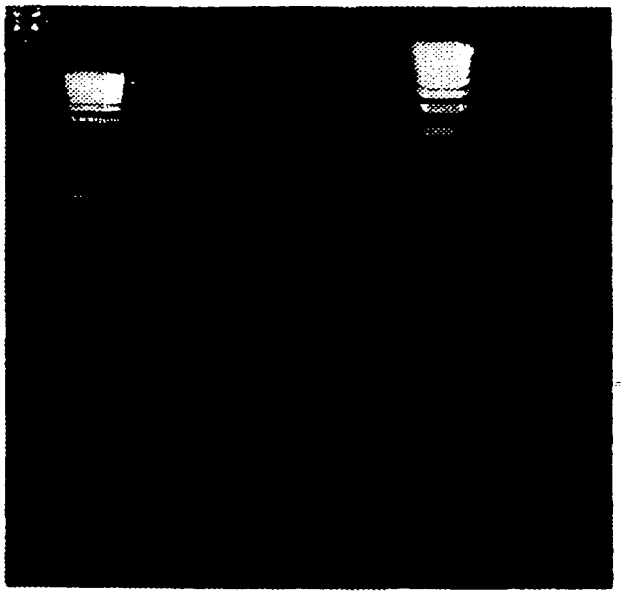


FIG.2A

FIG.2B

BNSDOCID: <WO 9717470A1>

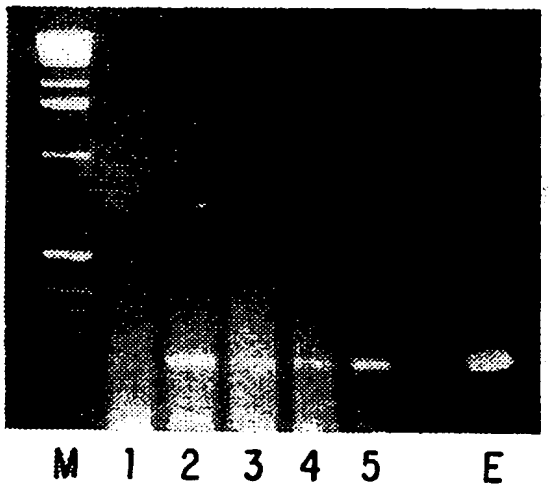


FIG.3A

FIG.3B

BNSDOCID: <WO 9717470A1>

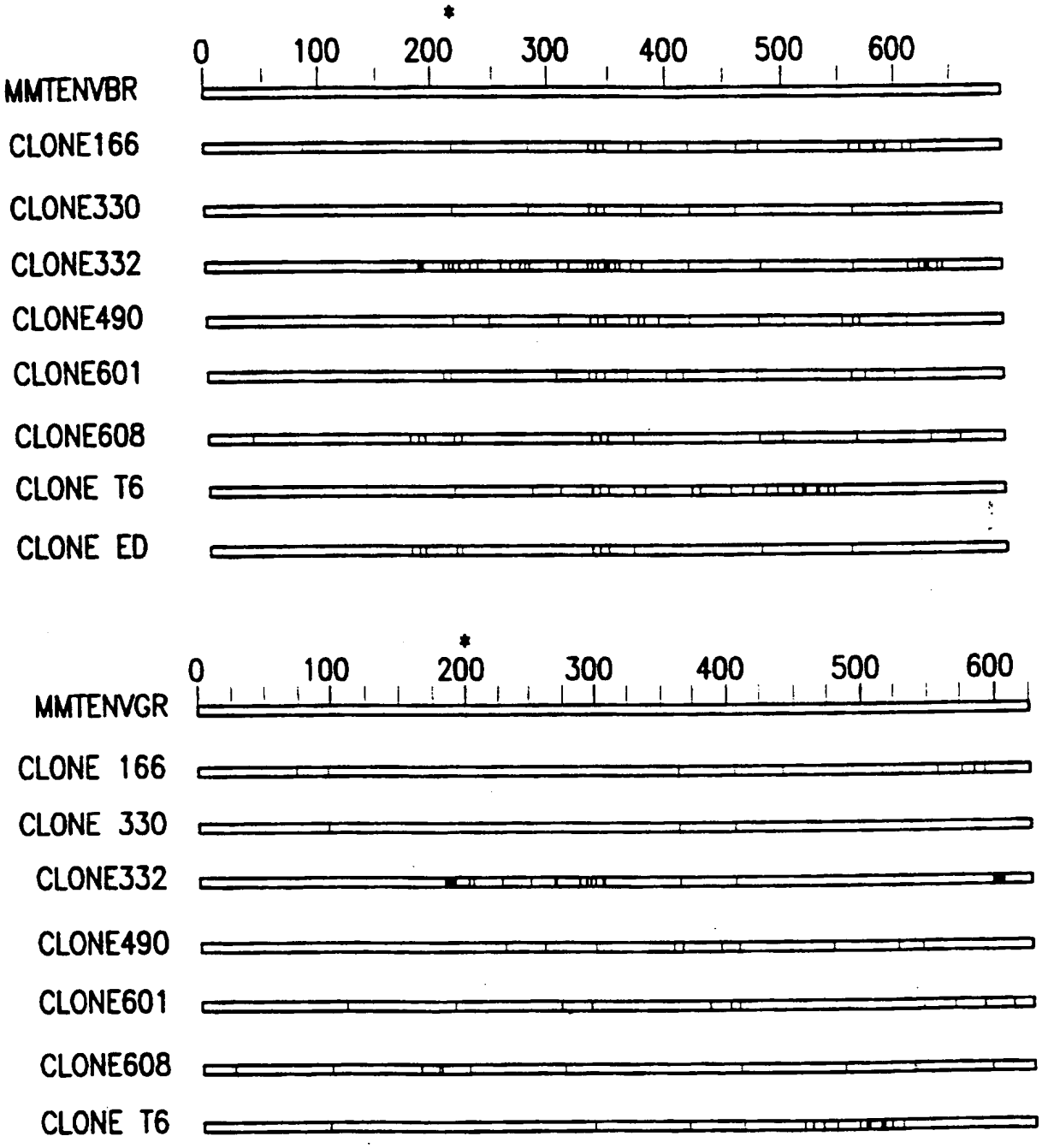


FIG.4

BNSDOCID: <WO 9717470A1>

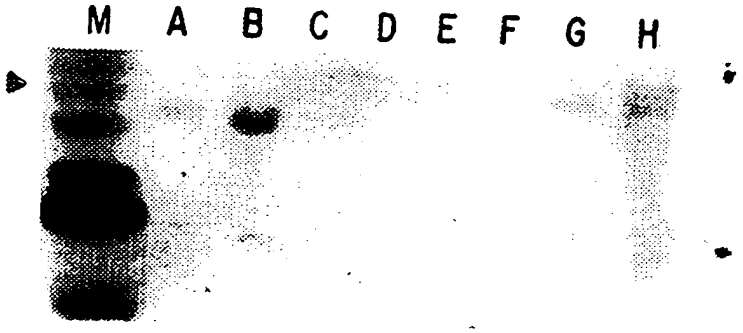


FIG.5

BNSDOCID: <WO 9717470A1>

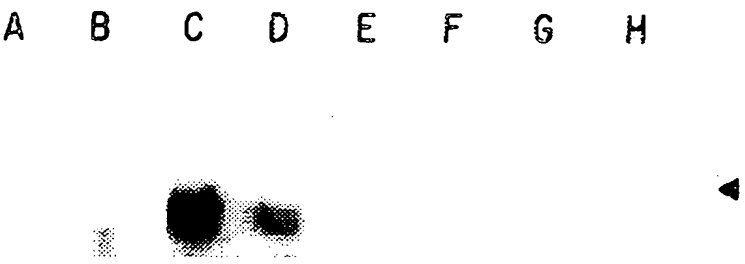


FIG.6

BNSDOCID: <WO 9717470A1>

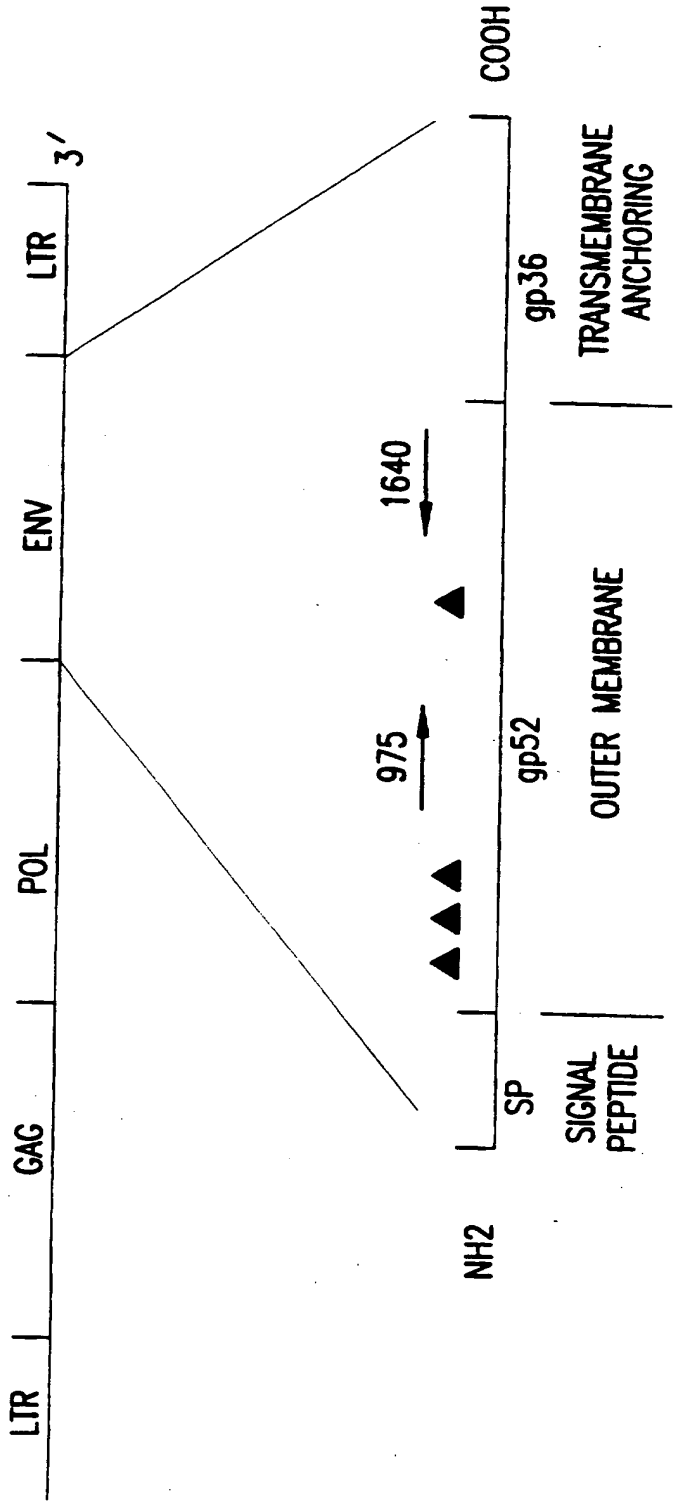


FIG. 7

BNSDOCID: <WO 9717470A1>

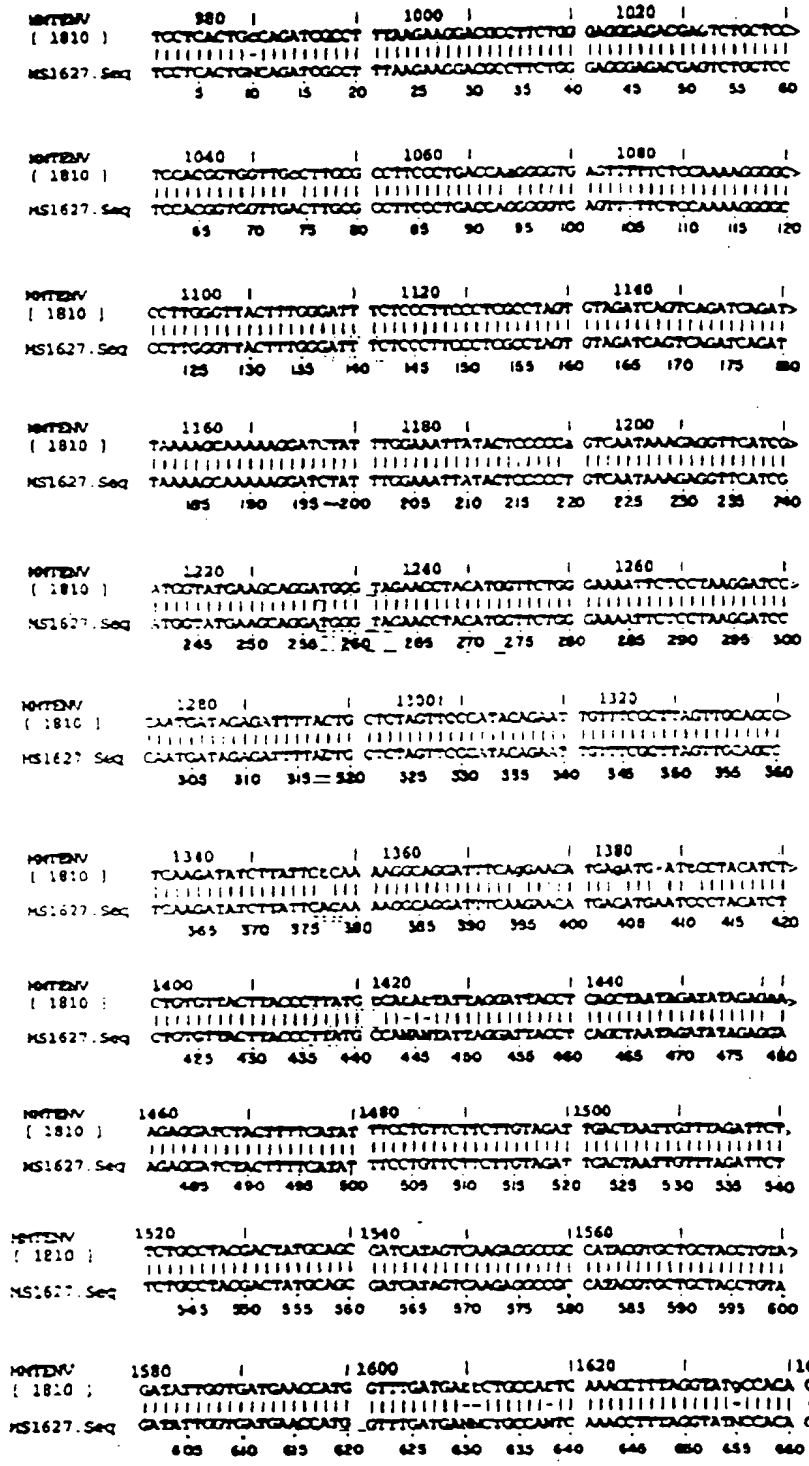


FIGURE 8

BNSDOCID: <WO 9717470A1>

CGAACAGACACAAACACACGAGAGGTGAATGTTAGGACTGTTGCAAGTTTA
CTCAAAAAACAGCACTCTTTTATATCATGGTTTACATAAGCATTTACATAAGA
CTTGGATAAGTTCCAAAAGAACATAGGAGAATAGAACACTCAGAGCTTAGAT
CAAAACATTTGATACCAAACCAAGTCAGGAAACCACTTGTCTCACATCCTTG
TTTTAAGAACAGTTTGTGACCCTGAACTTACTTAAACCTTGGGAACCGCAAN
GTTGGGCTCATAAAGGTTATCCATTATAGCTCATGCCAAAATTATCTGCAGA
AATGTGTTCCTAATTGTCTAGCCACTGCCCCCTCCCTTGGTATAATGAAAAT
CTTTCCCCCAACGTTCATCCCACTCCCCTAGATAAATATAATCATGTACCTGT
TGTTTTATGTCGTCTTTTTCTTCCTGAGTTAACACACACCAAGGAGGTCTAGC
TCTGGCGAGTCTTTCACGAAAGGGGAGGGATCTGTACAACACTTTATAGCC
GTTGACTGTGACCCACCTATCGAAATTTAAATCGTATCTTCCTGTATATGGTA
GCGGGGCGTCTGTTGGTCTGTAGATGTAAGTCCCGGTTGCCACCACCTGTC
TCCTATTTTGACAAGCGTACTCCTCTTTCCCCTTTTTAGTTCTAGGCCTGAGG
CCCTTAGTCCTTGCACCTGTTCTTCAACTGAGGTTGAGCGTCTCTTTCTATTT
TCTATTCCCATTTCTAACCTTTGAATTTGAGTAAATATAGTGCTAAAAGACAA
AGATTCATTTCTTAACATCATGATTAATAATCGACCTATTGGATTGGTCTTATT
GGTAAAAATATAATTTTTAGCAAGCATTCTTATTTCTATTTCTGAAGGACAAA
GTCGGTGTGGCTTGTAANAGGAANTTGGCTGTGGTCCTTGCCCCACGAGGA
AGGTCGAGTTCTCCGAATTGTTTAGATTGTAATCTTGCACAGAAGAGTTATTA
AAAGAATCAAGGGTGAGAGCCCTGCGAGCACGAACCGCAACTTCCCCCAAT
AGCCCCAGGCAAAGCAGAGCTATGCCAAGTTTGCAGCAGANAATGAGTATG
TCTTTGTCTGATGGGCTCATCCGCGTGCACGCAGACGGGTCGTCCTTGGTG
GGAAACAACCCCTTGGCTGCTTCTCTCCTAAGTGTAGGACACTCTCGGGAG
TTCAACCATTTCTGCTGCAGGCGCGGCATTTCCCCCTTTTTTCTTTTTTAAAA
GAAGCACGTTAAGATCTGACTGCACTTGGTCAAGGCTCTTCGCAAAGCACT
GGAAAATAACGGGGAAAATCATAAGTACTATGACCAAAAGCAGGGCTCCAA
CTCCTATAAAAATGAAATATTGTGTTCTAATCCAATGGATTTAAAGCCTTTAC
TCCATTGGCNAAGGANTGANCCAACCCCTGAGGTCCCTGCGTTCAAATTTTT
TTGCTCNTATCCTAATCCAATTGGTAACCCCGTTTNTTTTTGAAACTCATGTC
TTCAAATGCCCAATAAATGAGCCCTGGTTCTTTCCCAGCTCTCAGAAGCATT
ATACGGNANAGGTGTGACACAGCATAAAATCATAATTTGCATGACACCTAGT
GGACATTCTGGTCTTTAAGTTTGCCACATCTTGTCCCAACTCTAAAACTACTT
CTTCTAAAGCATTAAGTCTAGCTTTCAATTTTAAGTCTATTATTCTTTGTTCAG
ATNAGGCTAATGTAACATTTCTATGAAGATTATTAACAAACGTAGCAGTTTGC
ATCTCCTTAACTAAGGCAGTAGTAGCTACAGCAAAGGAAGTGATAATAGCAA
TTAAAGCAGATATGCCCAGAATAATGGCAGCGACGAATCGCTTAGCTCGAAT
TAAATCTGTGGCATACCTAAAGGTTTGAATGGCAGAATCATCAAACCATGGT
TCATCACCAATATCTACAGGTTACAACACATATGGCGGCCCCTTGAATATGA
ATCGCTGCATATCCGTNGGCAAAAAATCTAACCATTATTCCTCCTNCCNAAA
AACGGGATTTGAAANTTATNCCCCTTNCCCCNAACCCANACCGAGGTACCC
CATAATGNGGGGGGTATCTANAANAGGGCATAGGGGTAAGAAAAACGGCA
GAGNGGGATCNTTTATGTTCNGGAAATTCNGGGTTTGGGAGAATAAGATTCT
GGAGGCTGCAAATTAAGGGAAACATTNTGTATGGGGAATAGAGCAGTAAAA
TCTCTATCATGGGGATCTTTAGGGAGAATTTTCCCAGGAACCAAGTAGGTTC
NAACCCATCNTGCTTCATACCATCGATGAACNTCTTTATTGACAGGGGGAGT
ATAATTTCCAAATAGATCCTTTTTGTTTTTAATCTGATCTGACTGATCTACACT
AGGCGGGGGAAGGGAGAAATCCCAAAGTAACCCAAGGGCCCCTTTTGGAG
AAAAACTCACCCCCTGGTCAGGGAAGGCGCAAGGCAACCACCGTGGAGGA
GCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGGAGG
AGCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGG

FIG.9

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REDMOND et al. Sequence and expression of the mouse mammary tumour virus env gene. The EMBO Journal. 1983, Volume 2, Number 1, pages 125-131. See entire document.	1-20
A	FAFF et al., Retrovirus-like particles from the human T47D cell lines are related to mouse mammary tumour virus and are of human endogenous origin. Journal of General Virology. 21 May 1992, Volume 73, pages 1087-1097. See abstract.	1-20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 04 FEBRUARY 1997	Date of mailing of the international search report 18 MAR 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DIANNE REES JAB for Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CREPIN et al. Sequences Related to Mouse Mammary Tumor Virus Genome in Tumor Cells and Lymphocytes from Patients with Breast Cancer. Biochemical and Biophysical Research Communications. 13 January 1984, Volume 118, Number 1, pages 324-331. See entire document.	1-20
A	MESA-TEJADA et al. Detection in human breast carcinomas of an antigen immunologically related to a group-specific antigen of mouse mammary tumor virus. Proceedings of the National Academy of Sciences, USA. March 1978, Volume 75, Number 3, pages 1529-1533.	1-20

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12Q 1/68, 1/70; C12P 19/34; C07H 21/02, 21/04; G01N 33/53; C07K 15/28; 5/00

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, CANCERLIT, DGENE,DRUGU, EMBASES, MEDLINE, USPATFULL, TOXLIT, TOXLINE, JAPIO, WPIDS
search terms: MMTV, mouse mammary tumor virus, PCR, hybridization, antibodies, immunoassays, Westerns, searched SEQ. ID. Nos.

BNSDOCID: <WO 9717470A1>

THIS PAGE BLANK (USPTO)

CORRECTED VERSION*

CORRECTED VERSION**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, 1/70, C12P 19/34, C07H 21/02, 21/04, G01N 33/53, C07K 15/28, 5/00

A1

(11) International Publication Number: WO 97/17470

(43) International Publication Date: 15 May 1997 (15.05.97)

(21) International Application Number: PCT/US96/17877

(22) International Filing Date: 8 November 1996 (08.11.96)

(30) Priority Data:
08/555,394 9 November 1995 (09.11.95) US

(71)(72) Applicant and Inventor: HOLLAND, James, F. [US/US]; 31 Mamaroneck Road, Scarsdale, NY 10583 (US).

(72) Inventor: POGO, Beatriz, G., T.; 237 Nyac Avenue, Pelham, NY 10803 (US).

(74) Agents: CLARK, Richard, S. et al.; Brumbaugh, Graves, Donohue & Raymond, 30 Rockefeller Plaza, New York, NY 10112 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(57) Abstract

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

* (Referred to in PCT Gazette No. 31/1997, Section II)

** (Referred to in PCT Gazette No. 49/1997, Section II)

BNSDOCID: <WO 9717470A1>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Larvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

BNSDOCID: <WO 9717470A1>

Description

Detection Of Mammary Tumor Virus-Like Sequences In Human Breast Cancer

Cross-Reference to Related Application

This application is a continuation-in-part application of U.S. Serial No. 08/555,394, filed November 9, 1995.

Statement Regarding Federally Sponsored Research

This invention was made with funds from the U.S. government, which has certain rights in the invention.

Introduction

The present invention relates to materials and methods for diagnosing breast cancer in humans. It is based, at least in part, on the discovery that a substantial percentage of human breast cancer tissue samples contained nucleic acid sequences corresponding to a portion of the mouse mammary tumor virus env gene. In contrast, such sequences were absent in almost all other human tissues tested.

Background of the Invention

A large body of information has accumulated about the molecular biology of MMTV (reviewed in Slagle, B.L. et al., 1987, in "Cellular and Molecular Biology of Mammary Cancer", Kidwell et al., eds., Plenum Press, NY. pp 275-306). Mouse mammary tumor virus (MMTV) is associated with a high incidence of breast cancer in certain strains of mice (over 90% among females), and has been regarded as a potential model for human disease.

The MMTV virus does not carry a transforming oncogene, but rather acts as an insertional mutagen with several proviral insertion loci designated int-1

or wnt-1 (Nusse R. et al., 1982, Cell 31:99-109) int-2 (Peters, G. et al., 1983, Cell 33:369-377) int-3 (Gallahan, D. et al., 1987, J. Virol. 61:218-220) int-4 (Roelink, H. et al., 1990, Proc. Natl. acad. Sci. USA 87:4519-4523) and int-5 (Morris, V.L., et al. 1991, Oncogene Research 6:53-63), which encode for growth factors or other related proteins. These genes are not expressed in normal mammary tissue but become activated after integration of MMTV provirus into the adjacent chromosomal DNA.

The human homolog of the int-2 locus has been located on chromosome 11 (Casey, G. et al., 1986, Mol. Cell Biol. 6:502-510) and has been found amplified (in 15% of the breast cancers) and also expressed (Lidereau, R. et al., 1988, Oncogene Res 2:285-291; Zhou, D.J. et al., 1988, Oncogene 2:279-282; Liscia, D.S. et al., 1989, Oncogene 4:1219-1224; Meyers, S.L. et al., 1990, Cancer Res 50:5911-5918). It may be significant that in tumors from Parsi women, who have a high incidence of breast tumors, the int-2 locus is amplified in 50% of the cases (Barnabas-Sohi, N. et al., 1993, Breast Dis. 6:13-26). The amplification of int-2 and other genes in 11q13 is indicative of poor prognosis (Schuwring, E. et al., 1992, Cancer Research 52:5229-5234; Champeme, M-H, et al., 1995, Genes, Chromosomes and Cancer 12:128-133). Both mouse and human int-2 have been sequenced (Moore, R. et al., 1986, EMBO J 5:919-924). The gene encodes a protein of about 27 kilodaltons (KD) which shows homology to both basic and acidic fibroblast growth factors (Dickson, C. et al. 1987, Nature (London) 326:833).

However, efforts to demonstrate the presence of viruses in human breast cancer through search for viral particles, immunological cross-reactivity, or sequence homology have yielded contradictory results. Detectable MMTV env gene-related antigenic reactivity has been found in tissue sections of breast cancer

BNSDOCID: <WO 9717470A1>

(Mesa-Tejada et al., 1978, Proc. Natl. Acad. Sci. USA 75:1529-1533; Levine, P. et al., 1980, Proc. Am. Assoc. Cancer Res. 21:170; Lloyd, R. et al., 1983, Cancer 51:654-661), breast cancer cells in culture (Litvinov, S.V. and Golovkina, T.V., 1989, Acta Virologica 33:137-142), human milk (Zotter S. et al., 1980, Eur. J. Cancer 16:455-467) in sera of patients (Day, N.K. et al., 1981, Proc. Natl. Acad. Sci. USA 78:2483-2487), in cyst fluid (Witkin, S.S. et al., 1981, J. Clin. Invest. 67:216-222) and in particles produced by a human breast carcinoma cell line (Keydar, I. et al., 1984, Proc. Natl. Acad. Sci. USA 81:4188-4192). Sequence homology to MMTV has been found in human DNA under low stringency conditions of hybridization (Callahan, R. et al., 1982, Proc. Natl. Acad. Sci. USA 79:5503-5507) and RNA related to MMTV has been detected in human breast cancer cells (Axel, R. et al., 1972, Nature 235:32-36). The presence of MMTV related sequences in lymphocytes from patients with breast cancer has been reported (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331), as well as detection of reverse transcriptase (RT) activity in their monocytes (Al-Sumidaie, A.M. et al., 1988, Lancet 1:5-8). May and Westley (May and Westley, 1989, Cancer Research 49:3879-3883) have reported the presence of MMTV-like sequences arranged as tandem repeats only in DNA from breast cancer cells.

These results have been difficult to interpret, and theories linking MMTV or a related virus with human breast cancer have fallen out of favor, in view of the relatively recent discovery of human endogenous retroviral sequences ("HERs"; Westley, B. et al., 1986, J. Virol. 60:743-749; Ono, M. et al., 1986, J. Virol. 60:589-598; Faff, O. et al., 1992, J. Gen. Virology 73:1087-1097). Data which could be interpreted to demonstrate the presence of MMTV-related sequences could be more readily explained by endogenous human

BNSDOCID: <WO 9717470A1>

retroviral sequences. Adding further confusion to the picture, env-gene related antigenicity has been detected in epitopes of human proteins (Hareuveni, M. et al., 1990, Int. J. Cancer 46:1134-1135).

Brief Summary of the Invention

The present invention relates to methods for diagnosing breast cancer in humans in which the presence of mouse mammary tumor virus env gene-like sequences bears a positive correlation to the existence of malignant breast disease. It is based, at least in part, on the discovery that 38 to 40 percent of human breast cancer tissue samples tested contained gene sequences homologous to the mouse mammary tumor virus env gene that are substantially absent from other human tumors and tissues. The invention also relates to methods for diagnosing breast caner in humans in which the presence of retrovirus proviral fragments substantially homologous to the env gene and/or 3' LTR sequence of MMTV are detected. The molecular probes used in these experiments were designed to avoid cross-hybridization with endogenous human retroviral sequences. The present invention further provides for compositions of molecular probes which may be utilized in such diagnostic methods.

Brief Description of the Figures

FIGURE 1: Amplification of 660 bp of MMTV-like env gene. DNA was extracted from frozen tissues. PCR was performed using primers 1 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'^{32}P-end-labeled probe 2. Lanes 1 and 3: breast cancer; lanes 2 and 4: normal breast; lane 5: control reaction (no DNA); lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 510 bp band.

FIGURE 2: Nested PCR. A: 2% agarose gel electrophoresis. 1: Amplification of 686 bp of MMTV-like env

BNSDOCID: <WO 9717470A1>

gene sequences using primers 1 and 4 and the product of reaction A 1 as template. 2: Amplification of 250 bp of MMTV-like env gene sequences using primers 2 and 3. B, 1 and 2: Southern blot hybridization of the amplified products using probe 5'-^{32}P end-labeled probe 2a.

FIGURE 3: Amplification of 250 bp of MMTV-like env gene. DNA was extracted from paraffin-embedded tissue sections. PCR was performed using primers 2 and 3. A: 2% agarose gel electrophoresis. B: Southern blot hybridization using 5'-^{32}P-labeled probe 2a. Lane 1: normal breast; lanes 2 to 5: breast cancer; lane E: MMTV env gene. M: molecular weight marker. Arrow indicates 298 bp band.

FIGURE 4: Nucleotide sequence of the cloned MMTV env gene-like sequences as compared to the env sequences of the GR and BR6 strains of MMTV using the GCG program. *:potential glycosylation site, |:mismatch to MMTV.

FIGURE 5: Southern blot hybridization of genomic DNA. DNA was extracted from frozen tissues or cell lines, digested with EcoR1 and transferred to nitrocellulose paper. Hybridization with ^{32}P-labeled clone 166. DNA from A, B, and G: env gene positive breast cancer; C and D: env negative breast cancer; E and F: normal breast; H:MCF-7 cells. M: molecular weight marker, Arrow indicates 9kb band.

FIGURE 6: Southern blot hybridization of genomic DNA. Experimental conditions as in Fig. 5. DNA from A and B: env negative breast cancer; C and D: env positive breast cancer; E: molecular weight marker (non-labelled); F. to H: normal breast. Arrow indicates position of 9 kb marker.

FIGURE 7: Map of MMTV.

FIGURE 8: Comparison of the nucleic acid sequence of mouse mammary tumor env gene ("MMTENV"), showing residues 976-1640, with the nucleic acid sequence of a

BNSDOCID: <WO 9717470A1>

representative 660 bp sequence obtained by PCR reaction of DNA from human breast cancer tissue ("MS1627").

FIGURE 9: Sequence of an about 2.6 kb MMTV-like fragment detected in a human breast carcinoma.

Detailed Description of the Invention

The present invention relates to methods and compositions for diagnosing breast cancer in humans.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which (i) hybridizes to a gene of mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. A "gene of mouse mammary tumor virus" includes, but is not limited to, the gag, pol, and env genes and the 5′ LTR and 3′ LTR sequences of MMTV. In preferred embodiments of the invention, the mouse mammary tumor virus (hereafter "MMTV") gene is the env gene and/or the 3′ LTR sequence. The term "hybridize" is used to refer to routine DNA-DNA or DNA-RNA hybridization techniques under what would be regarded, by the skilled artisan, as stringent hybridization conditions. The phrase "is present" indicates that a native form of the molecule, in an unpurified state (for example, as part of chromosomal DNA), may be detected by a standard laboratory technique, such as Southern blot or polymerase chain reaction (PCR). To be "present", the molecule may be detectable by one technique but not others. To be present in "less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects", all non-breast cancer tissue samples are considered together, but the total number of samples must be large enough to give the 5 percent

BNSDOCID: <WO 9717470A1>

value statistical significance that would be reasonable to the skilled artisan.

In order to identify such a nucleic acid molecule, the sequence of MMTV may be compared, using a computer database, to known human DNA sequences, and portions of MMTV which are less than or equal to 25 percent homologous to a human sequence may be selected for further study. The term "homologous", as used herein, refers to the presence of identical residues; for example, a first sequence is considered 25 percent homologous to a second sequence if it shares 25 percent of the residues of the first sequence. Since there is relatively greater likelihood that MMTV may bear similarity to human retroviral-like sequences, it may be preferable to evaluate whether a particular MMTV nucleic acid sequence is homologous to such sequences, for example, as endogenous human retrovirus sequences. A prototype of such viruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598).

Once an MMTV gene sequence which is less than or equal to 25 percent homologous to a human DNA sequence, such as a human endogenous retroviral sequence, is identified, the presence of nucleic acid molecules having the MMTV gene sequence in human breast cancer tissues and other tissues may be evaluated. Such evaluations may be performed either by Southern blot techniques, or, preferably, by polymerase chain reaction (PCR) techniques, which are more sensitive. In such a way, MMTV gene sequences which (i) hybridize to at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects and (ii) hybridize to less than 5 percent of DNA samples prepared from human tissues other than breast cancer tissues may be identified. A nucleic acid molecule having a MMTV gene sequence which satisfies these requirements may then be used in diagnostic methods which detect the presence of such sequence in human

BNSDOCID: <WO 9717470A1>

breast tissue by standard techniques, including PCR techniques which assay for the presence of the molecule, but also, where appropriate, Southern blot, Northern blot, or Western blot techniques, to name but a few.

In preferred embodiments, the present invention relates to a portion of MMTV localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; see Fig. 7). This about 660 bp sequence (hereafter, "the 660 bp sequence") has been found to exhibit low (16 percent) homology to the prototype human endogenous retrovirus HERV-K10, using the IBI/Pustell Sequence Analysis Program, and has also been shown to be present in 121 (38.5%) of 314 unselected breast cancer tissue samples, in cultured breast cancer cells, in 2 of 29 breast fibroadenomas (6.9%) and in 2 of 107 breast specimens from reduction mammoplasties (1.8%). The sequence was not found in normal tissues including breast, lymphocytes from breast cancer patients nor in other human cancers or cell lines (see example section, infra). Similarly, an about 250 bp sequence (hereafter "the 250 bp sequence"), between positions 1388 and 1640 in the env gene, and therefore falling within the 660 bp sequence, was detected in 60 (39.7%) of 151 breast cancer, and in one of 27 normal breast samples assayed from paraffin-embedded sections. Cloning and sequencing of the 660 bp and 250 bp sequences demonstrated that they are 95-99% homologous to MMTV env gene, but not to the known human endogenous retroviruses ("HERs") nor to other viral or human genes (<18%).

In another preferred embodiment, the present invention relates to a a nucleic acid molecule which corresponds to a retroviral genomic fragment which has substantial homology to 3' LTR and/or env gene of the MMTV genome, and is found in a substantial percentage of breast cancer samples. By substantial percentage is

BNSDOCID: <WO 9717470A1>

meant at least 20% of tested breast cancer samples. Such a sequence is preferably comprised of the 3' LTR region and all or part of the env gene, although it may include more sequences of a retroviral genome. Most preferably, the sequence is at least comprised of an about 2.6 kb fragment which comprises the 1,228 base pair (bp) sequence of the 3' LTR sequence and 1,336 bp of the env gene sequence of MMTV (Fig. 9) (SEQ ID NO:20). When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

Retrovirus proviral sequences can be detected by PCR technology using primers derived from the MMTV genome. Such primers include primer 5L, containing the nucleotides 7376-7395 of the MMTV BR6 genome (5'-3': CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3, containing nucleotides 9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Other primers which correspond to or are homologous to MMTV sequences can be used as primers. Nucleotide fragments which correspond to or are homologous to the retroviral sequences isolated from the breast cancer samples can also be used to amplify additional retroviral fragments from the samples. Long PCR techniques can be used to amplify longer stretches of a proviral sequence.

The present invention provides for compositions comprising an isolated and purified nucleic acid molecule which hybridizes to the about 2.6 kb retroviral fragment shown in Fig. 9 under stringent conditions or is at least 90 percent homologous to said fragment using the MacVector homology determining program which may be used to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods.

BNSDOCID: <WO 9717470A1>

Nucleic acids having the 660 bp sequence, the 250 bp sequence, or all or part of the about 2.6 kb sequence, may therefore be used, according to the invention, to diagnose breast cancer in a subject, using methods which include PCR and Southern blot methods. Where PCR methods are used, primers such as those listed in Table 1, below, may be utilized.

The present invention provides for compositions comprising essentially purified and isolated nucleic acid having the 660 bp sequence or the 250 bp sequence or an at least five bp, and preferably greater than or equal to ten bp, subsequence thereof. In order to maintain the desired specificity, such nucleic acid molecules may preferably contain sequence falling within the 660 bp sequence, but preferably do not contain sequences from other portions of the MMTV genome, which may, undesirably, hybridize to human sequences which are not breast cancer specific, such as HERs. Accordingly, the present invention provides for compositions wherein the isolated and purified nucleic acid molecule comprises at least a portion having a nucleic acid sequence which hybridizes to a region of the mouse mammary tumor virus env gene between residues 976 and 1640, or between residues 1388 and 1640, and wherein the isolated and purified nucleic acid molecule does not hybridize to any other region of the MMTV genome.

The 660 bp sequence, in various embodiments, may have a number of nucleotide sequences. For example, in one embodiment, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MMTENV-like sequence" (SEQ ID NO:17), which depicts the MMTV env sequence between residues 976 and 1640. In a second series of embodiments, the 660 bp sequence may have a sequence as set forth in Fig. 8 and designated "MS1627" (SEQ ID NO:18), which depicts a predominant sequence for the 660 bp sequence as it has been defined by

BNSDOCID: <WO 9717470A1>

sequencing analysis of the products of PCR reactions using DNA from human breast cancer tissues. In still further embodiments, the 660 bp sequence may have various other nucleotide sequences obtained by sequencing the results of PCR reactions to detect the presence of 660 bp sequence in human breast cancer tissues.

In related embodiments, the present invention provides for compositions comprising PCR primers that may be used to detect the presence of the forementioned molecules or other MMTV-like sequences. For example, the compositions may comprise one or more of the following primer molecules (5′ - 3′): CCTCACTGCCAGATC (SEQ ID NO:1); GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2); CCTCACTGCCAGATCGCCT (SEQ ID NO:3); TACATCTGCCTGTGTTAC (SEQ ID NO:4); CCTACATCTGCCTGTGTTAC (SEQ ID NO:5); CCGCCATACGTGCTG (SEQ ID NO:6); ATCTGTGGCATACCT (SEQ ID NO:7); GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8); ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); TACAGGTAGCAGCACGTATG (SEQ ID NO:12); CGAACAGACACAAACACACG (SEQ ID NO:19).

The use of such compositions and molecules in PCR and Southern blot techniques is illustrated in the non-limiting examples set forth below. The correlation between the presence of the MMTV-related nucleic acid molecules described above and breast cancer allows such molecules and compositions to be utilized in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer, wherein the detection of such nucleic acid molecules bears a positive correlation to the existence of breast cancer in a human. The results of such evaluation, together with additional clinical symptoms, signs, and laboratory test values, may be used to formulate the complete diagnosis of the patient.

BNSDOCID: <WO 9717470A1>

In further related embodiments, the present invention provides for an essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects. In preferred embodiments, the MMTV gene is the env gene.

Such peptides may be used in the diagnosis of breast cancer. Accordingly, the present invention provides for a method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

The present invention also provides for antibodies (including monoclonal and polyclonal) antibodies which specifically bind to such peptides. Such antibodies may be used in methods of diagnosing breast cancer, for example, but not by way of limitation, by Western blot, immunofluorescent techniques, and so forth.

In nonlimiting embodiments of the invention, the skilled artisan may evaluate MMTV-like nucleic acid molecules for regions which would be considered likely to encode immunogenic peptides (using, for example, hydropathy plots). Such peptides may then be sequenced and used to produce antibodies that may be employed in diagnostic methods as set forth above.

For example, certain peptides encoded by portions of the 660 bp sequence have been synthesized. These

BNSDOCID: <WO 9717470A1>

peptides, which have the sequences LKRPGFQEHEMI (SEQ ID NO:13) and GLPHLIDIEKRG (SEQ ID NO:14), have been used to produce antibodies in rabbits, and the resulting antisera have successfully identified breast cancer cells positive for MMTV env-like sequences by PCR assay. Other peptides encoded by 660 bp sequence which may be useful according to the invention include TNCLDSSAYDTA (SEQ ID NO:15) and DIGDEPWFDD (SEQ ID NO:16).

6. Example: The Detection of Mouse Mammary Tumor Virus Env Gene-Like Sequences in Human Breast Cancer Cells and Tissues

6.1. Materials and Methods

DNA from breast cancer tissue and other human cancer tissues, human placentas, normal human tissues including breast, and from several human cell lines (including eight breast cancer cell lines), and two normal breast cell lines was extracted following the procedure of Delli Bovi et al. (1986, Cancer Res. 46:6333-6338). The DNA was resuspended in a solution containing 0.05 M Tris HCl buffer, pH 7.8, and 0.1 mM EDTA, and the amount of DNA recovered was determined by microfluorometry using Hoechst 33258 dye (Cesarone, C. et al., 1979, Anal Biochem 100:188-197). Plasmids containing the cloned genes of MMTV were obtained from the ATCC, propagated in Escherichia coli cultures and purified using anion-exchange minicolumns (Qiagen) or by precipitation with polyethylene glycol (Sambrook J., et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). Oligonucleotide primers were synthesized at the core facilities of the Brookdale Molecular Biology Center at Mount Sinai School of Medicine.

Polymerase chain reaction (PCR) was performed using Taq polymerase following the conditions recommended by the manufacturer (Perkin Elmer Cetus)

with regard to buffer, Mg^{2+} and nucleotide concentrations. Thermocycling was performed in a DNA cycler by denaturation at 94° C for 3 min. followed by either 35 or 50 cycles of 94°C for 1.5 min., 50° C for 2 min. and 72°C for 3 min. The ability of the PCR to amplify the selected regions of the MMTV env gene was tested by using as positive templates the cloned MMTV env gene and the genomic DNA of the MCF-7 cell line, since it was shown to express gp52 immunological determinants (Yang, N.S., et al., 1975, J. Natl. Cancer Inst. 61:1205-1208). Optimal Mg^{2+}, primer concentrations and requirements for the different cycling temperatures were determined with these templates. The master mix as recommended by the manufacturer was used. To detect possible contamination of the master mix components, a reaction without template was routinely tested. γ DNA and control primers provided by the manufacturer were used as control for polymerase activity. As an internal control, amplification of a 120 bp sequence estrogen receptor gene was assayed using primers designed and generously provided by Dr. Beth Schachter, (Mount Sinai School of Medicine, N.Y.). In addition, primers for actin 5 gene amplification were also used.

The product of the PCR was analyzed by electrophoresis in a 2% agarose gel. A 1 kb DNA ladder (Gibco BRL) was used to identify the size of the PCR product. To determine if the amplified sequences of the middle region of the 660 bp faithfully reproduced the sequences of the env gene of MMTV, an 18-mer sequence within the env gene was used as a probe for the 660 bp amplified sequence. The 18-mer probe was 5′ end-labeled with ^{32}P-ATP using T4 polynucleotide kinase and purified by the NENSORB nucleic acid purification cartridge (NEN). Southern blot hybridization was performed using the conditions described by (Saiki et al.,1985, Science 230:1350-1354).

BNSDOCID: <WO 9717470A1>

The product of the PCR (660 bp or 250 bp) was cloned directly from the reaction mixture into the TA cloning vector (Invitrogen) using the TA cloning kit and following the conditions recommended by the supplier. Direct cloning of the fragment isolated from the gel, was also performed. Plasmid DNA was purified by CsCl density gradient centrifugation or by precipitation with polyethylene glycol (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor), restricted with HindIII and EcoRl, electrophoresed in 2% agarose gels and transferred to nitrocellulose filters. Southern blot hybridization was carried out using a 5'-terminal labeled internal probe as described above. Cloning procedures were performed in laboratories totally separate from those where PCR was carried out. Automated DNA sequencing (using Applied Technology Sequencer Model 373A) was performed in the Brookdale Molecular Biology Center. Sequence homology was determined using the IBI MacVector GenBank and GCG Programs.

To prevent contamination of the samples, processing of human tissues was performed in a laminar flow hood. DNA extractions were done in a chemical hood located in a different room from that where PCR was performed. PCR assays were assembled in a biological hood provided with ultraviolet light. Aerosol resistant tips and dedicated positive-displacement pipettes were used throughout. All equipment used for PCR (microcentrifuge, electrophoresis apparatus, pipettors) was cleaned each time with 10% sodium hypochlorite to assure DNA decontamination (Prince and Andrus, 1992, Biotechniques 12:358-36). After the initial experiments were performed, the plasmid containing the MMTV <u>env</u> gene was frozen and never used again, to avoid contamination. However, to detect plasmid contamination from our own <u>env</u> gene clones,

primers were designed to amplify plasmid sequences. All the authentic MMTV env positive samples were then tested and found negative for plasmid contamination.

Southern blotting and hybridization were performed as described (Southern, E.M., 1975, J. Mol. Biol. 98:503-517), using the 660 bp cloned sequences labeled by the random primer procedure (Feinberg, A.P., et al., 1983, Anal. Biochem. 132:6-13). Prehybridization and hybridization were performed in a solution containing 6 x SSPE, 5% Denhardt's, 0.5% SDS, 50% formamide, 100 µg/ml denaturated salmon testis DNA, incubated for 18 hrs at 42°C, followed by washings with 2 x SSC and 0.5% SDS at room temperature and at 37°C and finally in 0.1 x SSC with 0.5% SDS at 68°C for 30 min (Sambrook et al., 1989, in "Molecular Cloning/A Laboratory Manual", Cold Spring Harbor). For paraffin-embedded tissue sections the conditions described by Wright and Manos (1990, in "PCR Protocols", Innis et al., eds., Academic Press, pp. 153-158) were followed using primers designed to detect a 250 bp sequence.

6.2. Results

6.2.1. Selection of Specific MMTV Env Gene Sequences

A computer search for MMTV env gene homologous sequences was first performed, since sequence homology between the human endogenous retroviral sequences and MMTV had been described. The prototype of this group of human endogenous retroviruses is HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598). The sequences of the env gene of MMTV (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504) were aligned with sequences of the env gene of the human endogenous retrovirus HERV-K10 (Ono, M. et al., 1986, J. Virol. 60:589-598), using the IBI/Pustell Sequence Analysis Program. A region of 660 bp of low homology (16%) was localized between MMTV env gene sequences 976 and 1640 (Majors, I.E. and Varmus, H.E., 1983, J Virol 47:495-504). This

BNSDOCID: <WO 9717470A1>

internal domain of the outer membrane of the env gene has only one glycosylation site and is highly conserved between strains. Two primers comprising 15 bp sequences at positions 976-990 (primer 1) and 1626-1640 (primer 3) were first synthesized. Later longer primers were synthesized (1N and 3N). An 18-mer sequence in the middle of the 660 bp MMTV env region (1388-1405) (primer 2) was used as a probe to identify the 660 bp sequence. A second oligomer probe was synthesized comprising the sequence 1554 to 1568 (primer 2a) to be used for hybridization when a sequence of around 250 bp (between positions 1388 and 1640) was amplified. For nested PCR reactions (Mullis, K.B. and Faloona, F.A., 1987, Meth Enzymol 155:335-350), another primer comprising sequences 1647 to 1661 (primer 4) was synthesized to be used with primer 1 in the first reaction and primers 2 and 3 in the second. Modified primers with GC clamps and extra sequences were also synthesized and used in the PCR (primers 1a and 3a). Another set of primers comprising sequences 974 to 1003 (5L) and 1558 to 1577 (3L) were subsequently developed because their Tm's matched and provided better amplification than the original primers. The sequences are represented in Table 1. All of them were productive in amplification reactions.

BNSDOCID: <WO 9717470A1>

Table 1. Primer and probe sequences and location in mouse mammary tumor virus env gene

Designation	Sequence (5'-3')	Location
1	CCTCACTGCCAGATC	976-990
1a	GGGAATTCCTCACTGCCAGATC	976-990
1N	CCTCACTGCCAGATCGCCT	976-993
2	TACATCTGCCTGTGTTAC	1388-1405
2N	CCTACATCTGCCTGTGTTAC	1386-1405
2a	CCGCCATACGTGCTG	1554-1568
3	ATCTGTGGCATACCT	1640-1626
3a	GGGAATTCATCTGTGGCATACCT	1640-1626
3N	ATCTGTGGCATACCTAAAGG	1640-1621
4	GAATCGCTTGGCTCG	1661-1647
5L	CCAGATCGCCTTTAAGAAGG	984-1003
3L	TACAGGTAGCAGCACGTATG	1558-1577

6.2.2. Detection of MMTV-Like Env Gene Sequences in Human Breast Tumor DNA

PCR was performed on DNA extracted from breast cancer tissues, normal breast tissues and from the plasmid containing the env gene of MMTV, using primers 1 and 3. Photographs of the ethidium bromide stained gels of the PCR product reveal the presence of an approximately 660 bp sequence in some of the tumors, (Fig. 1A, lanes 1 and 3) but not in the normal tissue samples (Fig. 1A, lanes 2 and 4). As a positive control the MMTV env gene was also amplified (Fig. 1A, lane E). Similar results were obtained with modified primers 1a, 3a, 3L and 5L. Southern blot hybridization of the gel with ^{32}P-labeled 18-mer oligonucleotide (primer 2) indicated that this internal sequence was present in the amplified material (Fig. 1B) and that the bands in the gel were not artifactual.

Our initial effort was to analyze a representative sample of breast cancer specimens as well as normal

BNSDOCID: <WO 9717470A1>

tissues and other tumors. To date 343 breast tumors have been processed, DNA extracted and PCR preformed. Of these 343 tumors, 314 were carcinomas and 29 were fibroadenomas. Amplification of sequences of 660 bp was observed in 121 of the carcinomas (38.5%) and in 2 of the 29 fibroadenomas (6.9%). These sequences were confirmed to be MMTV env gene-like sequences by hybridization with the labeled specific probe containing the internal sequences. These sequences were not detected in the DNAs extracted from 20 normal organs, 23 cancers from other organs and 26 samples of blood lymphocytes including 7 from breast cancer patients whose breast specimens were positive. From 107 samples of normal breast obtained from reduction mammoplasties, 2 were positive (1.8%). In addition to DNA from lymphocytes from seven positive patients, DNA from their normal breast tissue of the operated breast was tested in 4 cases. All were negative (Table 2). Finally, DNA of the MCF-7, and ED (a cell line developed in our laboratory from the pleural effusion of a patient with an env -positive breast tumor) breast cancer cell lines were shown to contain the 660 bp MMTV env gene-like sequences (Table 3), while four other breast cancer cell lines were positive only for the 250 bp sequence (T47-D, BT-474, BT-20 and MDA-MB-231).

BNSDOCID: <WO 9717470A1>

Tabl 2. Detection of MMTV env gene-like s quences in human DNA extract d from fresh or frozen tissues

Sample	Number	MMTV env gene sequences	% Positive
Breast Carcinomas	314	121	38.5%
Breast Fibroadenomas	29	2	6.9%
Normal Breasts	107	2	1.8%
*Normal Breasts	4	negative	
Tumors other than breast	23	negative	
Normal tissues	20	negative	
Lymphocytes	26	negative	
**Lymphocytes	7	negative	

* Histologically normal tissue from same breast as positive cancer.

** Lymphocytes from breast cancer patients who were positive for MMTV env gene sequences in the tumor.

BNSDOCID: <WO 9717470A1>

Table 3. Detection of MMTV env gene-like sequences in DNA from human cell lines in culture

Human Cell Lines		MMTV env gene sequence
MC-7	(breast carcinoma)	positive
T47-D	" "	negative
BT-20	" "	negative
MDA-MB-231	" "	negative
ZR-75-1	" "	negative
SK-BR 3	" "	negative
BT474	" "	negative
ED	" "	positive
MCF-10	(normal breast)	negative
HB-447	" "	negative
HL-60	(promyelocytic leukemia)	negative
K562	(erythroleukemia)	negative
Jurkat	(T cell leukemia)	negative
Hep 6-2	(hepatoma)	negative

The nested polymerase reaction was used in several instances to increase sensitivity and specificity, thus reducing the probability of false positives. In Fig. 2, results of a representative nested reaction are shown using primers 1 and 4 in the first reaction (Fig. 2A) and 2 and 3 for the 2nd reaction. The specificity of the reaction can be seen in the 2nd amplification (Fig. 2B).

To study a large number of samples and to be able to perform archival studies, PCR of paraffin-embedded tissue sections was also carried out. Primers 2 and 3 were used to amplify a 250 bp sequence within the 660 bp stretch when DNA was extracted from paraffin-embedded tissue sections since larger size sequences are difficult to amplify after fixation. Tumor DNA was amplified (Fig. 3A, lanes 2-5) whereas normal breast DNA was not (Fig. 3A, lane 1). The identification of

BNSDOCID: <WO 9717470A1>

this 250 bp sequence with the MMTV-like env gene was confirmed by hybridization with an internal probe (primer 2a) as shown in Fig. 3B. Using this procedure we have analyzed 151 breast cancer samples and found that 60 (39.7%) possess the 250 bp sequence. Of the 27 normal breast samples obtained from reduction mammoplasties assayed by this procedure, one was positive (3.7%). These results, in conjunction with those obtained from lymphocytes and from normal breast tissue of patients whose breast cancer was PCR positive, indicate that MMTV-like sequences are present in a significant number of human breast cancer DNA which cannot be explained by DNA polymorphism.

6.2.3. Cloning and Sequencing of the MMTV-Like Env Gene Sequences

To find out whether there was homology to MMTV env gene throughout the whole 660 bp stretch, the product of the PCR from 8 different tumors was cloned and sequenced. In Fig. 4 the sequence of different clones comprising around 600 bp are represented, as aligned to the MMTV env gene sequence of the GR and BR6 strains (Redmon, S. and Dickson, C., 1983, EMBO J. 2:125-131). This domain of the env gene in the GR strain is 100% homologous to the C_3H strain and 98% to the BR6 strain (Majors, I.E. and Varmus, H.E., 1983, J. Virol. 47:495-504; Moore, R. et al., 1987, J. Virol. 61:480-490). Evaluation of the clones indicated that homology to MMTV env gene varied from 95% to 99%. Another seven clones comprising only 250 bp were also sequenced. Homology to MMTV env gene varied from 95% to 99% (data not shown). When compared to the human endogenous provirus HERV-K10, the homology of all the clones was less than 15%. When compared against all known viral and human genes (more than 130,000 entries) using the 1B1 MacVector GenBank and GCG programs, the highest homology recorded was 18%.

6.2.4. Southern Blot Analysis Using Cloned Sequences

To investigate whether the env gene-like sequences were present in human DNA, Southern blot hybridization was performed using the cloned sequence as probe. DNAs from normal breast tissues, env positive or negative breast tumors, tumors other than breast and breast cancer cell lines were restricted with EcoRI and in some instances with PstI, BglII or KpnI. EcoRI is a frequent cutter restriction enzyme that digests MMTV proviral DNA between env and pol genes. Four different cloned 660 bp sequences were used as probes after labeling with ^{32}P by random prime-labeling. Results of some of the Southern blot hybridization experiments are shown in Fig. 5. They reveal the presence of a labeled restriction fragment migrating at approximately 7-8 kb in breast cancer DNA, in ED and two fragments in MCF-7 cells. Different restriction patterns were observed with the other three enzymes. The 660 bp sequence was absent in 10 normal tissues, 10 fibroadenomas and 10 tumors from other tissues. It is important to emphasize that hybridization conditions for these experiments were stringent (as described in Section 6.1) to avoid interference with endogenous sequences that might interact with the probes.

7. Example: Detection of a Retrovirus Proviral Fragment in Human Breast Cancer Cells and Tissues

7.1. Materials and Methods

To detect longer retrovirus proviral fragments in breast cancer samples, DNA was extracted from breast cancer carcinoma tissue samples as described above in Section 6.1. Two rounds of long PCR was performed on the DNA primers 5L (SEQ ID NO:11) and LTR3 (SEQ ID NO:19). The primer 5L contains nucleotides 7370-7395 of the MMTV BR6 genome (5′-3′: CCAGATCGCCTTTAAGAAGG) (SEQ ID NO:11) and primer LTR3 contains nucleotides

BNSDOCID: <WO 9717470A1>

9918-9927 of the MMTV BR6 genome (5'-3': CGAACAGACACAAAGCGACG) (SEQ ID NO:19). Long PCR was performed using protocols described by the manufacturer (Perkin Elmer, Foster City, CA). The amplified retroviral fragment isolated from the breast cancer sample was cloned into the TA cloning vector (Invitrogen) and automated sequencing was performed as described in Section 6.1.

7.2 Results

An approximately 2.6 kb retroviral fragment containing 1,228 bp of the 3' LTR sequence and 1,336 bp of the env gene sequence of a potential provirus was detected in a human breast carcinoma tissue sample by the long PCR technique using the 5L and LTR3 primers. The sequence of this retroviral fragment is shown in Fig. 9. (SEQ ID NO:20).

When compared with the two strains of MMTV C3H and BR6, the sequence homology was 90.8% and 90.7%, respectively, over the MMTV genomic fragment from nucleotides 7370-9937. When compared with the endogenous retroviral sequences (HUMERKA), sequence homology was only 58% in 36 bp and 71% in 74 bp.

8. Discussion

Search for virus-related sequences in human breast cancer has been hampered by great variation reported in previous studies, by the presence of endogenous retroviral sequences in human DNA and by the lack of sensitivity of the methods employed. The studies reported herein circumvent these deficiencies by focusing on sequences with low homology to human endogenous retroviruses, by investigating a large number of tumors and several types of controls and by using the most sensitive technology presently available.

The results indicate that unique MMTV env gene sequences were present in 38.5% of the breast cancer

BNSDOCID: <WO 9717470A1>

samples analyzed and 39.7% of archival samples of breast cancer and that these sequences were absent in normal tissues including lymphocytes from patients with positive breast cancer and in cancers other than breast. Normal breast tissue and fibroadenomas had a low frequency (1.8 to 6.9%) of positive results. When cloned and sequenced, the sequences were found to be highly homologous to MMTV env gene, but not to the endogenous retroviral sequences. Furthermore, experiments in which the cloned amplified sequences were used for hybridization with DNA from breast cancer or normal tissues revealed that homologous DNA was only present in breast cancer DNA. The results also indicate that a human breast carcinoma sample contained an about 2.6 kb MMTV-like fragment comprised of 1,336 bp of the env gene and 1,228 bp of the 3' LTR.

The detection of MMTV env gene sequences in two fibroadenomas out of 29 and in two normal breast tissue samples out of 107 samples is of uncertain significance. Although such results could potentially be artifactual, and thus may represent false positives, they may alternatively indicate the presence of histologically unrecognized cells that were or will be neoplastic.

Ninety percent (90%) of the breast cancers tested were invasive ductal carcinomas, which reflects the prevalence of this type of neoplasm. Most patients were node-positive which is probably artifactual since it was necessary that tumor size be sufficiently large to provide an aliquot for research and tumor size correlates with node positivity.

It is unlikely that differences in homology between MMTV env gene and the cloned human sequences are generated by errors committed by the Taq polymerase. It has been estimated that the rate of nucleotide misincorporation is 1×10^{-5} per cycle (Ehrlich et al, 1991, Science 252:1643-1651) and

therefore, only a total of 0.32 nucleotides misincorporated should be expected in 660 bp after 50 cycles. The differences in homology between clones from different patients is likely to represent heterogeneity of the env gene.

In contrast to earlier, ambiguous data associating MMTV-like sequences with human breast cancer, we have clearly demonstrated the existence of such sequences in breast cancer cells which cannot be explained by any known human endogenous retroviral sequence. Our data do not support the results of earlier studies which indicated that, as in the mouse, MMTV-like sequences were found in lymphocytes from two patients with breast cancer (Crepin, M. et al., 1984, Biochem. Biophys. Res. Comm. 118:324-331). The absence of MMTV env-like sequences in lymphocytes could reflect the fate of a unique lymphocyte subset over decades between initial encounter and the appearance of clinical breast cancer; alternatively, the human disease may differ from the mouse model. Results from attempts to identify unique MMTV-like pol gene sequences have shown that they cannot be distinguished from the reverse transcriptase sequences of endogenous retroviruses (Deen, K.C. and Sweet, R.W., 1986, J. Virol. 57:422-432).

The origin of the MMTV env gene-like and 3' LTR-like sequences found in tumor DNA could be the result of integrated MMTV-like sequences from a human mammary tumor virus. Polymorphism of endogenous retroviral sequences is conceivable but can be ruled out because these sequences were not detected in lymphocytes from the positive patients, in sections of the cancerous breast from which abnormal cells were absent, or in normal breast tissue from patients with MMTV env-like positive tumors. Recombination during tumorigenesis between endogenous sequences to resemble the MMTV env genes seems highly unlikely since no known gene or viral sequence is more than 18% homologous to the

BNSDOCID: <WO 9717470A1>

660 bp sequence. The longer about 2.6 kb MMTV-like fragment detected in a human breast carcinoma had minimal homology (58% in 36 bp nd 71% in 74 bp) to endogenous human retroviral sequences. Thus, the most conservative interpretation is that our findings represent exogenous sequences from an agent similar to MMTV. Recombination between endogenous and exogenous env gene sequences are known to accelerate the development of malignancies in mice (DiFronzo, N.L. and Holland, C.A., 1993, J. Virol. 67:3763-3770). Whether the MMTV-like sequences belong to an entire acquired provirus or to an exogenous fragment integrated into endogenous sequences, is presently not known. Experiments are in progress to distinguish between these possibilities.

Several genetic alterations have been identified in human breast cancer that can be useful as markers for prevention, detection or prognosis (reviewed in Runnenbaum, I. et al., 1991, Proc. Natl. Acad. Sci. USA 88:10657-10661). The BRCA1 and BRCA2 genes have recently been described. They account for at least 5% of breast cancer and are related to familial breast cancer (Miki, Y. et al., 1994, Science 266:66-71; Wooster, R. et al., 1994, Science 265:2088-2090). We have primary evidence that familial clustering of the MMTV env gene-like sequences occurs, accounting for an even higher percentage of cancers in affected families (Holland et al. 1994, Proc. Am. Assoc. Cancer Res 35:218). The presence of MMTV-like sequences may be correlated with special clinical disease status, may provide another potential molecular marker, and may distinguish a subset of human breast cancer for which viral etiology is tenable. This has implications for epidemiology, therapy and prevention.

Various publications are cited herein, the contents of which are hereby incorporated by reference in their entireties.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: HOLLAND, JAMES

(ii) TITLE OF THE INVENTION: DETECTION OF MAMMARY TUMOR VIRUS-LIKE SEQUENCES IN HUMAN BREAST CANCER

(iii) NUMBER OF SEQUENCES: 20

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
(B) STREET: 30 Rockefeller Plaza
(C) CITY: New York
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10112-0228

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: NOT YET ASSIGNED
(B) FILING DATE: 08-NOV-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER 08/555,394
(B) FILING DATE: 09-NOV-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Kole, Lisa B
(B) REGISTRATION NUMBER: 35,225
(C) REFERENCE/DOCKET NUMBER: 30363-PCT - 165/

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 212-408-2628
(B) TELEFAX: 212-765-2519
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCACTGCC AGATC 15

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGAATTCCT CACTGCCAGA TC 22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTCACTGCC AGATCGCCT 19

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TACATCTGCC TGTGTTAC 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

BNSDOCID: <WO 9717470A1>

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCTACATCTG CCTGTGTTAC 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCCATACG TGCTG 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTGTGGCA TACCT 15

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGAATTCAT CTGTGGCATA CCT 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCTGTGGCA TACCTAAAGG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATCGCTTG GCTCG 15

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCAGATCGCC TTTAAGAAGG 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAGGTAGC AGCACGTATG 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Lys Arg Pro Gly Phe Gln Glu His Glu Met Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Leu Pro His Leu Ile Asp Ile Glu Lys Arg Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

BNSDOCID: <WO 9717470A1>

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr Asn Cys Leu Asp Ser Ser Ala Tyr Asp Thr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: N-terminal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Asp Ile Gly Asp Glu Pro Trp Phe Asp Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCCTCACTGC CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGCCTTGCG CCTTCCCTGA CCAAGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCA GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCTCAA AAGGCAGGAT TTCAGGAACA TGAGATGATT CCTACATCTC 420
TGTGTTACTT ACCCTTATGT CATATTATTA GGATTACCTC AGCTAATAGA TATAGAGAAA 480
GAGGATCTAC TTTTCATATT TCCTGTTCTT CTTGTAGATT GACTAATTGT TTAGATTCTT 540
CTGCCTACGA CTATGCAGCG ATCATAGTCA AGAGGCCGCC ATACGTGCTG CTACCTGTAG 600
ATATTGGTGA TGAACCATGG TTTGATGATT CTGCCATTCA AACCTTTAGG TATGCCACAG 660
AT 662

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: lin ar

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCTCACTGN CAGATCGCCT TTAAGAAGGA CGCCTTCTGG GAGGGAGACG AGTCTGCTCC 60
TCCACGGTGG TTGACTTGCG CCTTCCCTGA CCAGGGGGTG AGTTTTTCTC CAAAAGGGGC 120
CCTTGGGTTA CTTTGGGATT TCTCCCTTCC CTCGCCTAGT GTAGATCAGT CAGATCAGAT 180
TAAAAGCAAA AAGGATCTAT TTGGAAATTA TACTCCCCCT GTCAATAAAG AGGTTCATCG 240
ATGGTATGAA GCAGGATGGG TAGAACCTAC ATGGTTCTGG GAAAATTCTC CTAAGGATCC 300
CAATGATAGA GATTTTACTG CTCTAGTTCC CATACAGAAT TGTTTCGCTT AGTTGCAGCC 360
TCAAGATATC TTATTCACAA AAGGCAGGAT TTCAAGAACA TGACATGAAT CCCTACATCT 420
CTGTGTTACT TACCCTTATG CCANANTATT AGGATTACCT CAGCTAATAG ATATAGAGGA 480
AGAGGATCTA CTTTTCATAT TTCCTGTTCT TCTTGTAGAT TGACTAATTG TTTAGATTCT 540
TCTGCCTACG ACTATGCAGC GATCATAGTC AAGAGGCCGC CATACGTGCT GCTACCTGTA 600
GATATTGGTG ATGAACCATG GTTTGATGAN NCTGCCANTC AAACCTTTAG GTATNCCACA 660
GAT 663

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGAACAGACA CAAACACACG 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGAACAGACA CAAACACACG AGAGGTGAAT GTTAGGACTG TTGCAAGTTT ACTCAAAAAA 60
CAGCACTCTT TTATATCATG GTTTACATAA GCATTTACAT AAGACTTGGA TAAGTTCCAA 120
AAGAACATAG GAGAATAGAA CACTCAGAGC TTAGATCAAA ACATTTGATA CCAAACCAAG 180
TCAGGAAACC ACTTGTCTCA CATCCTTGTT TTAAGAACAG TTTGTGACCC TGAACTTACT 240
TAAACCTTGG GAACCGCAAN GTTGGGCTCA TAAAGGTTAT CCATTATAGC TCATGCCAAA 300
ATTATCTGCA GAAATGTGTT CCTAATTGTC TAGCCACTGC CCCCTCCCTT GGTATAATGA 360
AAATCTTTCC CCCAACGTTC ATCCCACTCC CCTAGATAAA TATAATCATG TACCTGTTGT 420
TTTATGTCGT CTTTTTCTTC CTGAGTTAAC ACACACCAAG GAGGTCTAGC TCTGGCGAGT 480
CTTTCACGAA AGGGGAGGGA TCTGTACAAC ACTTTATAGC CGTTGACTGT GACCCACCTA 540
TCGAAATTTA AATCGTATCT TCCTGTATAT GGTAGCGGGG CGTCTGTTGG TCTGTAGATG 600
TAAGTCCCGG TTGCCACCAC CTGTCTCCTA TTTTGACAAG CGTACTCCTC TTTCCCCTTT 660
TTACTTCTAG GCCTGAGGCC CTTAGTCCTT GCACCTGTTC TTCAACTGAG GTTGAGCGTC 720
TCTTTCTATT TTCTATTCCC ATTTCTAACC TTTGAATTTG AGTAAATATA GTGCTAAAAG 780
ACAAAGATTC ATTTCTTAAC ATCATGATTA ATAATCGACC TATTGGATTG GTCTTATTGG 840
TAAAAATATA ATTTTTAGCA AGCATTCTTA TTTCTATTTC TGAAGGACAA AGTCGGTGTG 900
GCTTGTAANA GGAANTTGGC TGTGGTCCTT GCCCCACGAG GAAGGTCGAG TTCTCCGAAT 960
TGTTTAGATT GTAATCTTGC ACAGAAGAGT TATTAAAAGA ATCAAGGGTG AGAGCCCTGC 1020
GAGCACGAAC CGCAACTTCC CCCAATAGCC CCAGGCAAAG CAGAGCTATG CCAAGTTTGC 1080
AGCAGANAAT GAGTATGTCT TTGTCTGATG GGCTCATCCG CGTGCACGCA GACGGGTCGT 1140
CCTTGGTGGG AAACAACCCC TTGGCTGCTT CTCTCCTAAG TGTAGGACAC TCTCGGGAGT 1200
TCAACCATTT CTGCTGCAGG CGCGGCATTT CCCCCTTTTT TCTTTTTTAA AAGAAGCACG 1260
TTAAGATCTG ACTGCACTTG GTCAAGGCTC TTCGCAAAGC ACTGGAAAAT AACGGGGAAA 1320
ATCATAAGTA CTATGACCAA AAGCAGGGCT CCAACTCCTA TAAAAATGAA ATATTGTGTT 1380
CTAATCCAAT GGATTTAAAG CCTTTACTCC ATTGGCNAAG GANTGANCCA ACCCCTGAGG 1440
TCCCTGCGTT CAAATTTTTT TGCTCNTATC CTAATCCAAT TGGTAACCCC GTTTNTTTTT 1500
GAAACTCATG TCTTCAAATG CCCAATAAAT GAGCCCTGGT TCTTTCCCAG CTCTCAGAAG 1560
CATTATACGG NANAGGTGTG ACACAGCATA AAATCATAAT TTGCATGACA CCTAGTGGAC 1620
ATTCTGGTCT TTAAGTTTGC CACATCTTGT CCCAACTCTA AAACTACTTC TTCTAAAGCA 1680
TTAAGTCTAG CTTTCAATTT TAAGTCTATT ATTCTTTGTT CAGATNAGGC TAATGTAACA 1740
TTTCTATGAA GATTATTAAC AAACGTAGCA GTTTGCATCT CCTTAACTAA GGCAGTAGTA 1800
GCTACAGCAA AGGAAGTGAT AATAGCAATT AAAGCAGATA TGCCCAGAAT AATGGCAGCG 1860
ACGAATCGCT TAGCTCGAAT TAAATCTGTG GCATACCTAA AGGTTTGAAT GGCAGAATCA 1920
TCAAACCATG GTTCATCACC AATATCTACA GGTTACAACA CATATGGCGG CCCCTTGAAT 1980
ATGAATCGCT GCATATCCGT NGGCAAAAAA TCTAACCATT ATTCCTCCTN CCNAAAAACG 2040
GGATTTGAAA NTTATNCCCC TTNCCCCNAA CCCANACCGA GGTACCCCAT AATGNGGGGG 2100
GTATCTANAA NAGGGCATAG GGGTAAGAAA AACGGCAGAG NGGGATCNTT TATGTTCNGG 2160
AAATTCNGGG TTTGGGAGAA TAAGATTCTG GAGGCTGCAA ATTAAGGGAA ACATTNTGTA 2220
TGGGGAATAG AGCAGTAAAA TCTCTATCAT GGGGATCTTT AGGGAGAATT TTCCCAGGAA 2280
CCAAGTAGGT TCNAACCCAT CNTGCTTCAT ACCATCGATG AACNTCTTTA TTGACAGGGG 2340
GAGTATAATT TCCAAATAGA TCCTTTTTGT TTTTAATCTG ATCTGACTGA TCTACACTAG 2400
GCGGGGGAAG GGAGAAATCC CAAAGTAACC CAAGGGCCCC TTTTGGAGAA AAACTCACCC 2460
CCTGGTCAGG GAAGGCGCAA GGCAACCACC GTGGAGGAGC AGACTCGTCT CCCTCCCAGA 2520
AGGCGTCCTT CTTAAAGGCG ATCTGGAGGA GCAGACTCGT CTCCCTCCCA GAAGGCGTCC 2580
TTCTTAAAGG CGATCTGG 2598

Claims

1. A composition comprising an oligonucleotide primer which may be used to detect the presence of a nucleic acid molecule which (i) hybridizes to the env gene of a mouse mammary tumor virus; (ii) is present in at least 38 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 7 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

2. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATC (SEQ ID NO:1).

3. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCCTCACTGCCAGATC (SEQ ID NO:2).

4. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTCACTGCCAGATCGCCT (SEQ ID NO:3).

5. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence TACATCTGCCTGTGTTAC (SEQ ID NO:4).

6. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCTACATCTGCCTGTGTTAC (SEQ ID NO:5).

7. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCGCCATACGTGCTG (SEQ ID NO:6).

BNSDOCID: <WO 9717470A1>

8. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence ATCTGTGGCATACCT (SEQ ID NO:7).

9. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence GGGAATTCATCTGTGGCATACCT (SEQ ID NO:8).

10. The composition of claim 1, wherein the oligonucleotide primer comprises a sequence selected from the group consisting of ATCTGTGGCATACCTAAAGG (SEQ ID NO:9); GAATCGCTTGGCTCG (SEQ ID NO:10); CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11); and TACAGGTAGCAGCACGTATG (SEQ ID NO:12).

11. An essentially purified peptide encoded by a nucleic acid molecule which (i) hybridizes to a gene of MMTV; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

12. An antibody which specifically binds to the peptide of claim 11.

13. The peptide according to claim 11 which comprises the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13).

14. An antibody which specifically binds to the peptide of claim 13.

15. The peptide according to claim 11 which comprises the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14).

BNSDOCID: <WO 9717470A1>

16. A method of diagnosing breast cancer in a human subject, comprising detecting the presence of a peptide encoded by a nucleic acid molecule which (i) hybridizes to the env gene of 3' LTR of a mouse mammary tumor virus; (ii) is present in at least 20 percent of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) is present in less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

17. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence LKRPGFQEHEMI (SEQ ID NO:13) is detected by the binding of an antibody specific to the peptide.

18. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence GLPHLIDIEKRG (SEQ ID NO:14) is detected by the binding of an antibody specific to the peptide.

19. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence TNCLDSSAYDTA (SEQ ID NO:15) is detected by the binding of an antibody specific to the peptide.

20. The method according to claim 16, wherein the presence of a peptide comprising the amino acid sequence DIGDEPWFDD (SEQ ID NO:16) is detected by the binding of an antibody specific to the peptide.

21. A composition comprising an oligonucleotide primer which may be used to detect the presence of a

nucleic acid molecule which (i) hybridizes to a nucleic acid comprised of a sequence selected from the group consisting of the env gene and the 3' LTR of a mouse mammary tumor virus; (ii) is present in a substantial percentage of DNA samples prepared from breast cancer tissue of different human subjects; and (iii) hybridizes to less than 5 percent of DNA samples prepared from tissues other than breast cancer tissue from different human subjects.

22. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CCAGATCGCCTTTAAGAAGG (SEQ ID NO:11).

23. The composition of claim 1, wherein the oligonucleotide primer comprises the sequence CGAACAGACACAAACACACG (SEQ ID NO:19).

BNSDOCID: <WO 9717470A1>

FIG. 1A

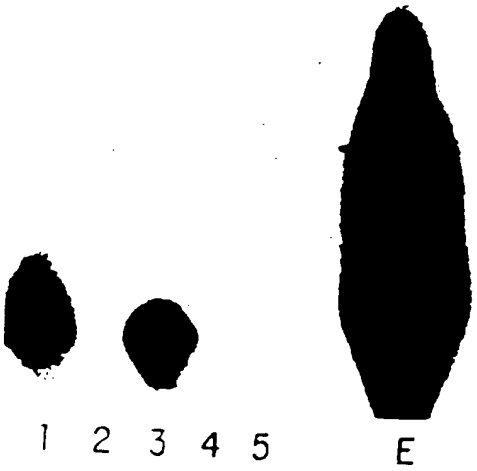


FIG. 1B

BNSDOCID: <WO 9717470A1>

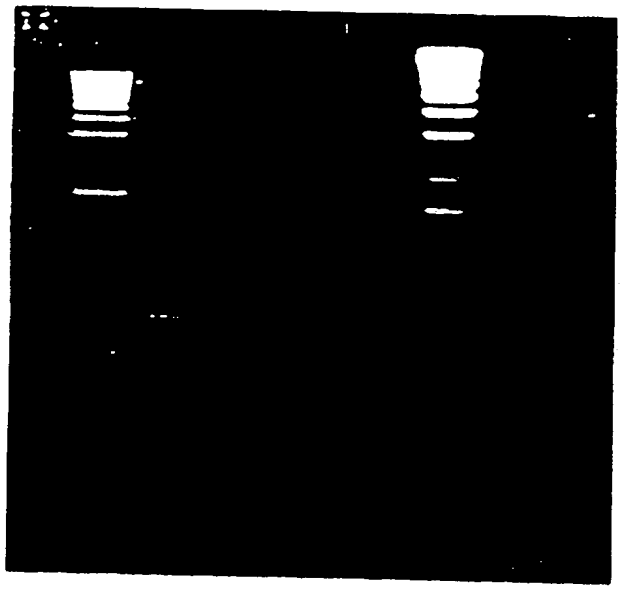


FIG.2A

FIG.2B

BNSDOCID: <WO 9717470A1>

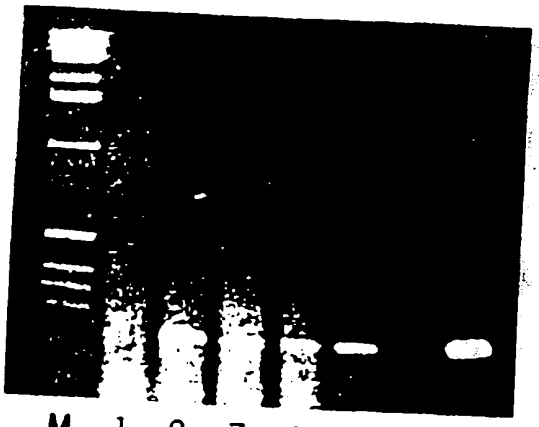


FIG.3A

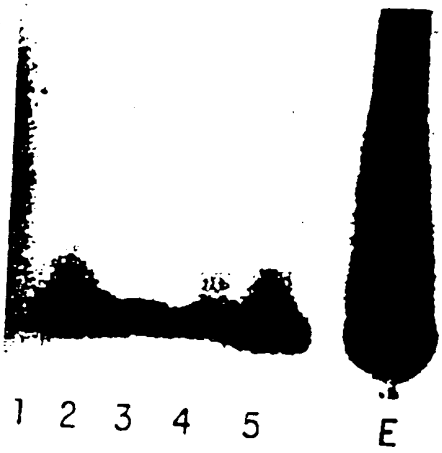


FIG.3B

BNSDOCID: <WO 9717470A1>

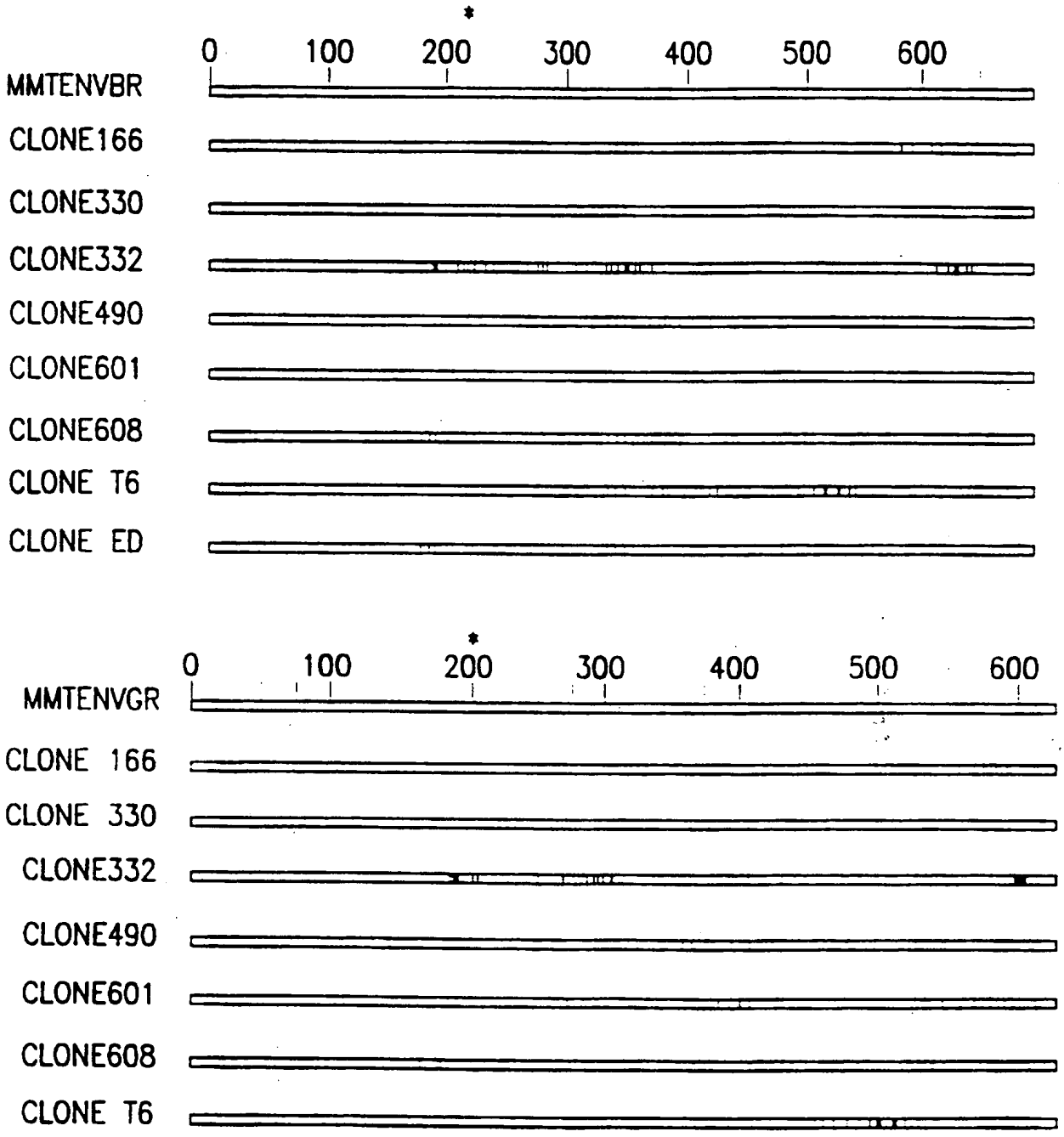


FIG.4

BNSDOCID: <WO 9717470A1>

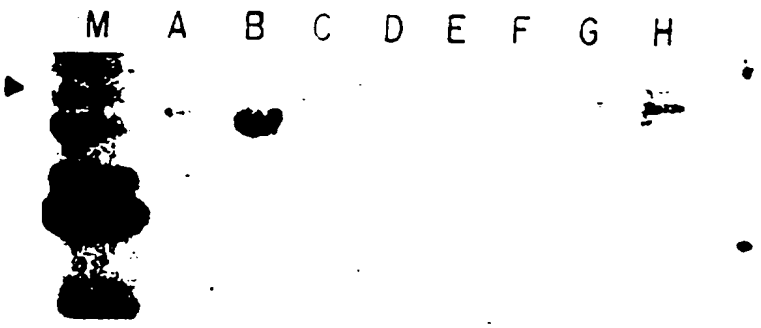


FIG. 5

BNSDOCID: <WO 9717470A1>

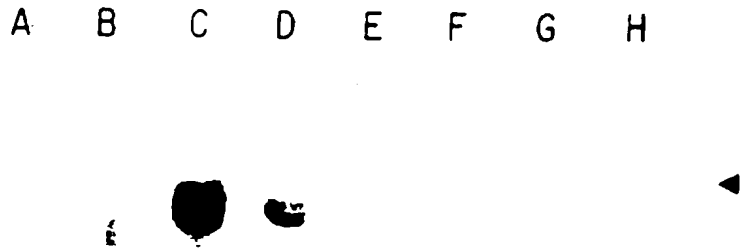


FIG.6

BNSDOCID: <WO 9717470A1>

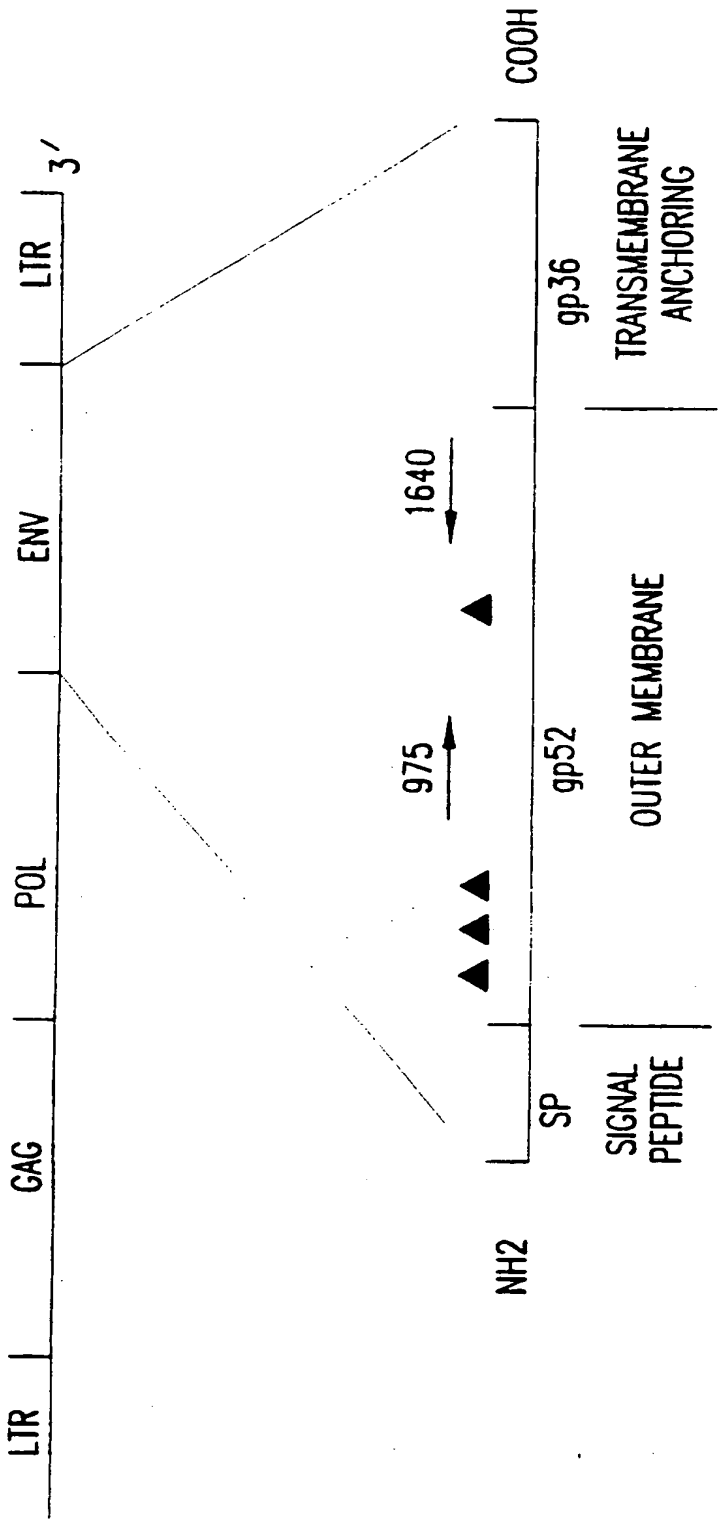


FIG. 7

BNSDOCID: <WO 9717470A1>

BNSDOCID: <WO 9717470A1>

SUBSTITUTE SHEET (RULE 26)

```
MMTENV 974 TCCTCACTGCCAGATCGCCTTTAAGAAGGACGCCTTCTGGGAGGGAGACG 1023
           |||||||||:||||||||||||||||||||||||||||||||||||||||
MS1627   1 TCCTCACTGNCAGATCGCCTTTAAGAAGGACGCCTTCTGGGAGGGAGACG 50

MMTENV1024 AGTCTGCTCCTCCACGGTGGTTGCCTTGCGCCTTCCCTGACCAAGGGGTG 1073
           |||||||||||||||||||| |||||||||||||||||| ||||||
MS1627  51 AGTCTGCTCCTCCACGGTGGTTGACTTGCGCCTTCCCTGACCAGGGGGTG 100

MMTENV1074 AGTTTTTCTCCAAAAGGGGCCCTTGGGTTACTTTGGGATTTCTCCCTTCC 1123
           |||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 101 AGTTTTTCTCCAAAAGGGGCCCTTGGGTTACTTTGGGATTTCTCCCTTCC 150

MMTENV1124 CTCGCCTAGTGTAGATCAGTCAGATCAGATTAAAAGCAAAAAGGATCTAT 1173
           |||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 151 CTCGCCTAGTGTAGATCAGTCAGATCAGATTAAAAGCAAAAAGGATCTAT 200

MMTENV1174 TTGGAAATTATACTCCCCCAGTCAATAAAGAGGTTCATCGATGGTATGAA 1223
           |||||||||||||||||| ||||||||||||||||||||||||||||||
MS1627 201 TTGGAAATTATACTCCCCCTGTCAATAAAGAGGTTCATCGATGGTATGAA 250
```

FIG. 8A

```
MMTENV1224 GCAGGATGGGTAGAACCTACATGGTTCTGGGAAAATTCTCCTAAGGATCC 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 251 GCAGGATGGGTAGAACCTACATGGTTCTGGGAAAATTCTCCTAAGGATCC 300

MMTENV1274 CAATGATAGAGATTTTACTGCTCTAGTTCCCATACAGAATTGTTTCGCT 1323
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 301 CAATGATAGAGATTTTACTGCTCTAGTTCCCATACAGAATTGTTTCGCT 350

MMTENV1324 TAGTTGCAGCCTCAAGATATCTTATTCTCAAAAGGCAGGATTTCAGGAA 1373
           |||||||||||||||||||||||||| |||||||||||||||| |||
MS1627 351 TAGTTGCAGCCTCAAGATATCTTATTCACAAAAGGCAGGATTTCAAGAA 400

MMTENV1374 CATGAGATG-ATCCCTACATCTCTGTGTTACTTACCCTTATGTCATATT 1423
           ||||| || ||||||||||||||||||||||||||||| ||-|-|
MS1627 401 CATGACATGAATCCCTACATCTCTGTGTTACTTACCCTTATGCCANANT 450

MMTENV1424 ATTAGGATTACCTCAGCTAATAGATATAGAGAAAGAGGATCTACTTTTC 1473
           |||||||||||||||||||||||||||| ||||||||||||||||||
MS1627 451 ATTAGGATTACCTCAGCTAATAGATATAGAGGAAGAGGATCTACTTTTC 500
```

FIG. 8B

```
MMTENV1474 ATATTTCCTGTTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCC 1523
           ||||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 501 ATATTTCCTGTTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCC 550

MMTENV1524 TACGACTATGCAGCGATCATAGTCAAGAGGCCGCCATACGTGCTGCTACC 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||
MS1627 551 TACGACTATGCAGCGATCATAGTCAAGAGGCCGCCATACGTGCTGCTACC 600

MMTENV1574 TGTAGATATTGGTGATGAACCATGGTTTGATGATTCTGCCATTCAAACCT 1623
           ||||||||||||||||||||||||||||||||||||-|||||||-||||||||||
MS1627 601 TGTAGATATTGGTGATGAACCATGGTTTGATGANNCTGCCANTCAAACCT 650

MMTENV1624 TTAGGTATGCCACA GAT 1640
           ||||||||-|||||| |||
MS1627 651 TTAGGTATNCCACA GAT 667
```

FIG. 8C

BNSDOCID: <WO 9717470A1>

CGAACAGACACAAACACACGAGAGGTGAATGTTAGGACTGTTGCAAGTTTA
CTCAAAAAACAGCACTCTTTTATATCATGGTTTACATAAGCATTTACATAAGA
CTTGGATAAGTTCCAAAAGAACATAGGAGAATAGAACACTCAGAGCTTAGAT
CAAAACATTTGATACCAAACCAAGTCAGGAAACCACTTGTCTCACATCCTTG
TTTTAAGAACAGTTTGTGACCCTGAACTTACTTAAACCTTGGGAACCGCAAN
GTTGGGCTCATAAAGGTTATCCATTATAGCTCATGCCAAAATTATCTGCAGA
AATGTGTTCCTAATTGTCTAGCCACTGCCCCCTCCCTTGGTATAATGAAAAT
CTTTCCCCCAACGTTCATCCCACTCCCCTAGATAAATATAATCATGTACCTGT
TGTTTTATGTCGTCTTTTTCTTCCTGAGTTAACACACACCAAGGAGGTCTAGC
TCTGGCGAGTCTTTCACGAAAGGGGAGGGATCTGTACAACACTTTATAGCC
GTTGACTGTGACCCACCTATCGAAATTTAAATCGTATCTTCCTGTATATGGTA
GCGGGGCGTCTGTTGGTCTGTAGATGTAAGTCCCGGTTGCCACCACCTGTC
TCCTATTTTGACAAGCGTACTCCTCTTTCCCCTTTTTAGTTCTAGGCCTGAGG
CCCTTAGTCCTTGCACCTGTTCTTCAACTGAGGTTGAGCGTCTCTTTCTATTT
TCTATTCCCATTTCTAACCTTTGAATTTGAGTAAATATAGTGCTAAAAGACAA
AGATTCATTTCTTAACATCATGATTAATAATCGACCTATTGGATTGGTCTTATT
GGTAAAAATATAATTTTTAGCAAGCATTCTTATTTCTATTTCTGAAGGACAAA
GTCGGTGTGGCTTGTAANAGGAANTTGGCTGTGGTCCTTGCCCCACGAGGA
AGGTCGAGTTCTCCGAATTGTTTAGATTGTAATCTTGCACAGAAGAGTTATTA
AAAGAATCAAGGGTGAGAGCCCTGCGAGCACGAACCGCAACTTCCCCCAAT
AGCCCCAGGCAAAGCAGAGCTATGCCAAGTTTGCAGCAGANAATGAGTATG
TCTTTGTCTGATGGGCTCATCCGCGTGCACGCAGACGGGTCGTCCTTGGTG
GGAAACAACCCCTTGGCTGCTTCTCTCCTAAGTGTAGGACACTCTCGGGAG
TTCAACCATTTCTGCTGCAGGCGCGGCATTTCCCCCTTTTTTCTTTTTTAAAA
GAAGCACGTTAAGATCTGACTGCACTTGGTCAAGGCTCTTCGCAAAGCACT
GGAAAATAACGGGGAAAATCATAAGTACTATGACCAAAAGCAGGGCTCCAA
CTCCTATAAAAATGAAATATTGTGTTCTAATCCAATGGATTTAAAGCCTTTAC
TCCATTGGCNAAGGANTGANCCAACCCCTGAGGTCCCTGCGTTCAAATTTTT
TTGCTCNTATCCTAATCCAATTGGTAACCCCGTTTNTTTTTGAAACTCATGTC
TTCAAATGCCCAATAAATGAGCCCTGGTTCTTTCCCAGCTCTCAGAAGCATT
ATACGGNANAGGTGTGACACAGCATAAAATCATAATTTGCATGACACCTAGT
GGACATTCTGGTCTTTAAGTTTGCCACATCTTGTCCCAACTCTAAAACTACTT
CTTCTAAAGCATTAAGTCTAGCTTTCAATTTTAAGTCTATTATTCTTTGTTCAG
ATNAGGCTAATGTAACATTTCTATGAAGATTATTAACAAACGTAGCAGTTTGC
ATCTCCTTAACTAAGGCAGTAGTAGCTACAGCAAAGGAAGTGATAATAGCAA
TTAAAGCAGATATGCCCAGAATAATGGCAGCGACGAATCGCTTAGCTCGAAT
TAAATCTGTGGCATACCTAAAGGTTTGAATGGCAGAATCATCAAACCATGGT
TCATCACCAATATCTACAGGTTACAACACATATGGCGGCCCCTTGAATATGA
ATCGCTGCATATCCGTNGGCAAAAAATCTAACCATTATTCCTCCTNCCNAAA
AACGGGATTTGAAANTTATNCCCCTTNCCCCNAACCCANACCGAGGTACCC
CATAATGNGGGGGGTATCTANAANAGGGCATAGGGGTAAGAAAAACGGCA
GAGNGGGATCNTTTATGTTCNGGAAATTCNGGGTTTGGGAGAATAAGATTCT
GGAGGCTGCAAATTAAGGGAAACATTNTGTATGGGGAATAGAGCAGTAAAA
TCTCTATCATGGGGATCTTTAGGGAGAATTTTCCCAGGAACCAAGTAGGTTC
NAACCCATCNTGCTTCATACCATCGATGAACNTCTTTATTGACAGGGGGAGT
ATAATTTCCAAATAGATCCTTTTTGTTTTTAATCTGATCTGACTGATCTACACT
AGGCGGGGGAAGGGAGAAATCCCAAAGTAACCCAAGGGCCCCTTTTGGAG
AAAAACTCACCCCCTGGTCAGGGAAGGCGCAAGGCAACCACCGTGGAGGA
GCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGGAGG
AGCAGACTCGTCTCCCTCCCAGAAGGCGTCCTTCTTAAAGGCGATCTGG

FIG.9

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 5, 91.2, 7.1, 7.2; 536 23.1, 24.3, 24.33; 530/388.1, 300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REDMOND et al. Sequence and expression of the mouse mammary tumour virus env gene. The EMBO Journal. 1983, Volume 2, Number 1, pages 125-131. See entire document.	1-20
A	FAFF et al., Retrovirus-like particles from the human T47D cell lines are related to mouse mammary tumour virus and are of human endogenous origin. Journal of General Virology. 21 May 1992, Volume 73, pages 1087-1097. See abstract.	1-20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 04 FEBRUARY 1997	Date of mailing of the international search report 18 MAR 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DIANNE REES JAB for Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CREPIN et al. Sequences Related to Mouse Mammary Tumor Virus Genome in Tumor Cells and Lymphocytes from Patients with Breast Cancer. Biochemical and Biophysical Research Communications. 13 January 1984, Volume 118, Number 1, pages 324-331. See entire document.	1-20
A	MESA-TEJADA et al. Detection in human breast carcinomas of an antigen immunologically related to a group-specific antigen of mouse mammary tumor virus. Proceedings of the National Academy of Sciences, USA. March 1978, Volume 75, Number 3, pages 1529-1533.	1-20

BNSDOCID: <WO 9717470A1>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17877

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12Q 1/68, 1/70; C12P 19/34; C07H 21/02, 21/04; G01N 33/53; C07K 15/28; 5/00

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, CANCERLIT, DGENE,DRUGU, EMBASES, MEDLINE, USPATFULL, TOXLIT, TOXLINE, JAPIO, WPIDS
search terms: MMTV, mouse mammary tumor virus, PCR, hybridization, antibodies, immunoassays, Westerns, searched SEQ. ID. Nos.

BNSDOCID: <WO 9717470A1>

THIS PAGE BLANK (USPTO)